



Fotis Jannidis (Hg.)

Digitale Literatur- wissenschaft

DFG-Symposion 2017

OPEN ACCESS



J.B. METZLER



Germanistische Symposien

Berichtsbände

Fotis Jannidis
(Hrsg.)

Digitale Literaturwissenschaft

DFG-Symposium 2017



J.B. METZLER

Hrsg.
Fotis Jannidis
Universität Würzburg
Würzburg, Deutschland



Gedruckt mit Unterstützung der Deutschen Forschungsgemeinschaft

ISSN 0936-3890 ISSN 2945-9494 (electronic)
Germanistische Symposien
ISBN 978-3-476-05885-0 ISBN 978-3-476-05886-7 (eBook)
<https://doi.org/10.1007/978-3-476-05886-7>

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

© Der/die Herausgeber bzw. der/die Autor(en) 2022. Dieses Buch ist eine Open-Access-Publikation.

Open Access Dieses Buch wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Buch enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag, noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Cover: Finken & Bumiller Stuttgart, Abbildung: hector-j-rivas/unsplash

Planung/Lektorat: Ferdinand Pöhlmann

J.B. Metzler ist ein Imprint der eingetragenen Gesellschaft Springer-Verlag GmbH, DE und ist ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Heidelberg Platz 3, 14197 Berlin, Germany

Inhaltsverzeichnis

Digitale Literaturwissenschaft. Zur Einführung	1
Fotis Jannidis	
Sektion I. Literatur und Literaturwissenschaft unter digitalen Bedingungen	
Einführung	19
Simone Winko	
Die Lesbarkeit der Welt. Digitalität als Herausforderung der Literaturwissenschaft	27
Michael Stolz	
Empirie – Beschreibung – Interpretation. Über den Platz von Computermodellen in den hermeneutisch-historisch orientierten Literaturwissenschaften	57
Jonas Kuhn	
Was verändert sich eigentlich? Korpusanalytisch basierte Wissenschaftsgeschichte der germanistischen Praxis am Beispiel der <i>Deutschen Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte</i>	97
Fotis Jannidis, Steffen Martus, Leonard Konle und Jörn Kreutel	
Lieder als kollektive Texte. Digitale Verschriftung und populäre Philologie auf <i>Genius</i>	141
Alexander Nebrig	
Tausend Tode, tausend Autoren, tausend Texte? Zur Textualität digitaler Literatur	173
Julia Nantke	
Wie und wozu analysiert man Blogs? Überlegungen zwischen <i>Digital Philology</i>, <i>Distant Reading</i> und Kulturpoetik	201
Jörg Schuster	

Diskussionsbericht Sektion I. Literatur(wissenschaft) unter digitalen Bedingungen	229
Jan Borkowski	
Sektion II. Digitale Edition und Annotation	
Einführung	243
Andrea Rapp	
Zwischenräume. Kommentierende Annotation und hermeneutische Bedeutungerschließung in digitalen Texten	249
Matthias Bauer, Gabriel Viehhauser und Angelika Zirker	
Unterwegs zum Text ohne Herausgeber und ohne Leser. Eine medienpragmatische und medientheoretische Standortbestimmung der digitalen Edition	281
Manuel Braun, Sonja Glauch und Florian Kragl	
Modellierung von Textgeschichte. Bedingungen digitaler Analyse und Schlussfolgerungen für die Editorik	307
Gerrit Brüning	
Digital Genetic Editions. Towards Macroanalysis across Versions	339
Dirk Van Hulle	
Annotation als Markup <i>avant la lettre</i>	353
Jan Christoph Meister	
Digitale Hermeneutik. Computergestütztes <i>Close Reading</i> als literaturwissenschaftliches Forschungsparadigma?	385
Evelyn Gius	
Historisch-kritische Ausgabe digital. Eine Reformulierung der neugermanistischen Edition	419
Rüdiger Nutt-Kofoth	
Diskussionsbericht Sektion II. Digitale Edition und Annotation	451
Luise Borek	
Sektion III. Textanalyse	
Einführung	467
Jan Christoph Meister	
What are they Talking About? A Systematic Exploration of Theme Identification Methods for Character Speech in Dramatic Texts	473
Nils Reiter und Marcus Willand	
Operationalising the Change. Dispersion of Polish literary life (1989–2002)	509
Maciej Maryl	

Quantitative Semantik. <i>Word Embedding Models</i> für literaturwissenschaftliche Fragestellungen	535
Christof Schöch	
<i>Small Worlds, Beat Charts</i> und die Netzwerkanalyse dramatischer Texte. Reflexionen aus dem <i>Rabbit Hole</i>	563
Peer Trilcke	
<i>Burrows's Delta</i> und Z-Score-Differenz im Netzwerkvergleich. Analysen zum <i>Deutschen Novellenschatz</i> von Paul Heyse und Hermann Kurz (1871–1876).	597
Thomas Weitin	
Operationalisierung der Metapher zur quantifizierenden Untersuchung deutschsprachiger literarischer Texte im Übergang vom Realismus zur Moderne	629
J. Berenike Herrmann	
Diskussionsbericht Sektion III. Textanalyse.	663
Jan Horstmann	
Sektion IV. Schnittstellen der digitalen Literaturwissenschaft	
Einführung	675
Thomas Stäcker	
Literaturwissenschaft und Bibliothek – Eine Beziehung im digitalen Wandel	679
Thomas Stäcker	
‘Could be the translation, of course’. Analysing the perception of literary fiction and literary translations	709
Karina van Dalen-Oskam	
Briefeditionen vernetzen.	729
Stefan Dumont	
Netzliteratur als Soziales Medium. Zur Kritik des Werkbegriffs	751
Thomas Ernst	
Ein Quantum Literatur. Empirische Daten zu einer Theorie des literarischen Textumfangs	777
Frank Fischer und Robert Jäschke	
Reassembling the Novel. The English Novel, 1789–1919.	813
Allen Riddell und Michael Betancourt	
Diskussionsbericht Sektion IV. Schnittstellen	833
Constanze Baum	

Herausgeber- und Autorenverzeichnis

Über den Herausgeber

Fotis Jannidis ist Professor für Computerphilologie und Neuere Deutsche Literaturgeschichte an der Julius-Maximilians-Universität Würzburg. Sein Hauptarbeitsgebiet sind die Computational Literary Studies, und er ist auch der Koordinator des gleichnamigen DFG-Schwerpunktprogramms 2019 ff.

Autorenverzeichnis

Matthias Bauer ist Professor für Englische Philologie (Literaturwissenschaft) an der Universität Tübingen und Sprecher des *Graduiertenkollegs 1808 Ambiguität: Produktion und Rezeption*. Seine Forschungsschwerpunkte liegen u. a. an der Schnittstelle von Literatur und Sprache und auf dem Gebiet des Textverstehens.

Constanze Baum forscht und lehrt als wissenschaftliche Mitarbeiterin am Institut für deutsche Literatur der Humboldt-Universität zu Berlin. Ihre Arbeitsschwerpunkte bilden u. a. Literatur und die Künste sowie *Digitales Publizieren*.

Michael Betancourt is the principal research scientist with Symplectomorphic, LLC, where he develops theoretical, methodological, and computational tools to support practical Bayesian inference, and a core developer of Stan.

Luise Borek ist wissenschaftliche Assistentin am Institut für Sprach- und Literaturwissenschaft der Technischen Universität Darmstadt. Sie forscht und unterrichtet in der germanistischen Mediävistik und der digitalen Philologie.

Jan Borkowski ist seit 2010 wissenschaftlicher Mitarbeiter am Seminar für Deutsche Philologie der Universität Göttingen. Sein literaturgeschichtlicher Arbeitsschwerpunkt liegt auf der deutschen Literatur um 1800 und 1900. Weitere Forschungsgebiete sind Literaturtheorie und Fragen der Interpretationsmethodik.

Manuel Braun ist Professor für Germanistische Mediävistik an der Universität Stuttgart. Derzeit arbeitet er vor allem auf den Gebieten der digitalen Editorik, der quantitativen Literaturwissenschaft sowie der mittelalterlichen Literaturästhetik.

Gerrit Brüning ist Referent für Forschung im *Stabsreferat Forschung/Kolleg Friedrich Nietzsche* der Klassik Stiftung Weimar. Er war am Freien Deutschen Hochstift, der Goethe-Universität Frankfurt sowie weiteren Universitäten tätig und ist seit 2009 an der Edition von Goethes *Faust* beteiligt.

Stefan Dumont ist wissenschaftlicher Mitarbeiter im Bereich *Digital Humanities* an der Berlin-Brandenburgischen Akademie der Wissenschaften. Seine Forschungsschwerpunkte sind Digitale Editionen (insbesondere von Briefen) und ihre Standardisierung und Vernetzung.

Thomas Ernst forscht und lehrt an der Universität Antwerpen im Bereich der Neueren Deutsche Literaturwissenschaft und der Netzliteraturwissenschaft. Mehr Informationen unter <https://www.thomasernst.net>.

Frank Fischer ist Professor für Digital Humanities an der FU Berlin. Sein Arbeitsschwerpunkt sind die Computational Literary Studies. Er ist u. a. Mitherausgeber des internationalen Dramenkorporus DraCor.

Evelyn Gius ist Professorin für Digital Philology – Neuere deutsche Literaturwissenschaft am Institut für Sprach- und Literaturwissenschaft der Technischen Universität Darmstadt. Ihre Forschungsschwerpunkte sind Computational Literary Studies, digitale Narratologie und der Einsatz manueller Annotation in der interdisziplinären und kollaborativen Textanalyse.

Sonja Glauch lehrt germanistische Mediävistik an der Friedrich-Alexander-Universität Erlangen-Nürnberg. Zu den Schwerpunkten ihrer Forschung gehören Literatur- und Erzähltheorie sowie Poetik, Medialität und Überlieferungsgeschichte der deutschen Literatur des Mittelalters.

J. Berenike Herrmann ist Professorin für Neuere Deutsche Literaturwissenschaft mit Schwerpunkt Literaturtheorie/Digital Humanities an der Universität Bielefeld. Zu ihren Forschungsschwerpunkten in den Computational Literary Studies gehören neben Methodologie und der Stilistik derzeit vor allem die Deutschschweizer Literaturgeschichte, fiktionale Repräsentationen von Landschaft und Emotion sowie literarische Wertung.

Jan Horstmann hat an der Universität Hamburg im Fach Germanistik über „Theaternarratologie“ promoviert und arbeitet seit 2016 im Bereich Digital Humanities. Von 2017 bis 2020 koordinierte er das DFG-Projekt *forTEXT* (<https://fortext.net>) zur Dissemination von DH-Methoden. Von 2020 bis 2021 war er Geschäftsführer des Forschungsverbunds Marbach Weimar Wolfenbüttel (<https://mww-forschung.de>) und leitete das an der Klassik Stiftung Weimar ansässige Digitale Labor mit dem Schwerpunkt der digitalen Sammlungsforschung. Seit Juli 2021 leitet er das Service Center for Digital Humanities an der Westfälischen Wilhelms-Universität Münster (<https://www.uni-muenster.de/SCDH/>).

Robert Jäschke ist Professor für Information Processing and Analytics am Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin. Er ist Mitbegründer des Social-Bookmarking-Systems *BibSonomy*.

Leonard Konle ist Doktorand der Digital Humanities an der Universität Würzburg. Seine Schwerpunkte sind Textkomplexität, Gattungsklassifikation und die Adaption neuronaler Netze für literaturwissenschaftliche Fragestellungen.

Florian Kragl ist Professor für Deutsche Philologie im europäischen Kontext an der Friedrich-Alexander-Universität Erlangen-Nürnberg. Im Zentrum seiner Forschung steht die deutsche Literatur des Mittelalters, auch unter vergleichendem Einschluss weiterer altgermanischer Dialekte und benachbarter Sprachen. Wichtige Arbeitsschwerpunkte sind (die digitale) Editionsphilologie, Poetik und Ästhetik sowie Erzähltheorie.

Jörn Kreutel ist Professor für Medieninformatik an der Berliner Hochschule für Technik. Seine Forschungsschwerpunkte sind die Modellierung und Analyse bibliographischer Datenbestände und die Konzeption nutzerfreundlicher Bedienoberflächen für Datenerfassung und -zugriff.

Jonas Kuhn Professor für Computerlinguistik an der Universität Stuttgart, arbeitet zu Algorithmen für die Sprach- und Textanalyse. Einer der Schwerpunkte liegt auf deren Anwendung in den Digital Humanities.

Steffen Martus Professor für Neuere deutsche Literatur an der Humboldt-Universität zu Berlin. 2015 Gottfried Wilhelm Leibniz-Preis der DFG. Mitherausgeber des „Jahrbuchs der Deutschen Schillergesellschaft“ sowie der „Studien zur deutschen Literatur“; Redaktionsleitung von „Text + Kritik“. Letzte Buchpublikationen: „Aufklärung. Das deutsche 18. Jahrhundert - ein Epochenbild“ (2015); „Geistesarbeit. Eine Praxeologie der Geisteswissenschaften“ (2022, gemeinsam mit Carlos Spoerhase).

Maciej Maryl is assistant professor and the founding head of the *Digital Humanities Centre at the Institute of Literary Research* of the Polish Academy of Sciences. He is interested in data-driven sociology of literature. Website: maryl.org.

Jan Christoph Meister ist Professor im Ruhestand an der Uni Hamburg und war dort bis 2020 Universitätsprofessor für *Digital Humanities* mit dem Schwerpunkt Neuere deutsche Literatur an der Universität Hamburg. Seine Forschungsschwerpunkte sind digitale Textanalyse und Textannotation, Narratologie und korpusgestützte Erzähltextanalyse.

Julia Nantke ist Juniorprofessorin in der Neueren deutschen Literaturwissenschaft mit dem Schwerpunkt Digital Humanities für Schriftartefakte an der Universität Hamburg. Forschungsschwerpunkte: digitale Literatur und digitale Literaturwissenschaften, Literaturtheorie, Editionswissenschaft, Materialität und Medialität der Literatur, Literatur und Kunst der Avantgarden.

Alexander Nebrig lehrt Neuere Deutsche Literaturwissenschaft an der Heinrich-Heine-Universität Düsseldorf. Aktuell beschäftigt er sich mit der Globalisierung der deutschsprachigen Literatur und der Literaturgeschichte des interlingualen Lizenzraumes.

Rüdiger Nutt-Kofoth ist wissenschaftlicher Mitarbeiter für Neuere deutsche Literatur sowie Editionswissenschaft an der Bergischen Universität Wuppertal und wissenschaftlicher Koordinator *des Graduiertenkolleg 2196 Dokument – Text – Edition* ebendort. Arbeitsschwerpunkte sind u. a. Editionswissenschaft, Wissenschaftsgeschichte und Literaturtheorie.

Andrea Rapp ist Professorin für Germanistik – Computerphilologie und Mediävistik an der Technischen Universität Darmstadt. Ihre Forschungen umfassen sowohl die Entwicklung digitaler Analysetechnologien, die Erstellung digitaler Editionen und Wörterbücher, aber auch den nachhaltigen Aufbau von Forschungsinfrastrukturen sowie die Reflexion von Digitalität in der philologischen Forschung und im Bereich des Kulturellen Erbes.

Nils Reiter ist Professor für Digital Humanities und Sprachliche Informationsverarbeitung an der Universität zu Köln. Mit einem Hintergrund in der Computerlinguistik forscht er zu Operationalisierung und Interpretierbarkeit in den Digital Humanities, wobei ein Schwerpunkt in der komputationellen Literaturwissenschaft liegt.

Allen Riddell is Assistant Professor in the *Luddy School of Informatics, Computing, and Engineering* at Indiana University Bloomington. His research explores applications of modern statistical methods in the humanities and allied social sciences. His research interests include sociology of literature, publishing history, library digitization, and text mining.

Christof Schöch ist Professor für Digital Humanities an der Universität Trier und Ko-Direktor des Kompetenzzentrums – *Trier Center for Digital Humanities (TCDH)*. Er arbeitet zu Ressourcen und Methoden für die Computational Literary Studies. Siehe auch: <http://www.christof-schoech.de>.

Jörg Schuster ist Wissenschaftlicher Mitarbeiter und Privatdozent am Institut für deutsche Literatur der Goethe-Universität Frankfurt und ist Mitherausgeber der Tagebücher von Harry Graf Kessler. Seine Arbeitsschwerpunkte sind Literatur, Ästhetik und Medien vom 18. Jahrhundert bis zur Gegenwart.

Thomas Stäcker ist Direktor der Universitäts- und Landesbibliothek Darmstadt und nebenamtlicher Professor für *Digital Humanities* an der Fachhochschule Potsdam. Seine Arbeits- und Forschungsschwerpunkte liegen im Bereich der Buch- und Bibliotheksgeschichte und den *Digital Humanities*, insb. *Digitale Editionen* und *Digitale Sammlungen*.

Michael Stolz Professor für germanistische Mediävistik an der Universität Bern, leitet die digitale Neuedition des in über 80 Handschriftenzeugen überlieferten *Parzival*-Romans Wolframs von Eschenbach (www.parzival.unibe.ch). Weitere

Arbeitsschwerpunkte sind mediävistische Komparatistik und historische Lese-forschung. (Gossembrot-Projekt: www.gossembrot.unibe.ch).

Peer Trilcke ist Professor für deutsche Literatur des 19. Jahrhunderts und Leiter des Theodor-Fontane-Archivs an der Universität Potsdam. Seine Forschungsschwerpunkte sind die Literatur und Kultur des 19. Jahrhunderts, Computational Literary Studies, Theorie und Praxis des Literaturarchivs sowie Digitales Kulturerbe.

Karina van Dalen-Oskam is head of the department of Literary Studies at Huygens Institute for the History and Culture of the Netherlands and professor in computational literary studies at the University of Amsterdam. Her research focuses on stylometry, with a special interest in proper names (comparative literary onomastics).

Dirk Van Hulle is Professor of Bibliography and Modern Book History at the University of Oxford, director of the Oxford Centre for Textual Editing and Theory (OCTET) and director of the Centre for Manuscript Genetics at the University of Antwerp. His research interests are genetic criticism, digital scholarly editing, and the work of twentieth-century authors such as Samuel Beckett and James Joyce.

Gabriel Viehhauser ist Professor für *Digital Humanities* an der Universität Stuttgart. Seine Forschungsschwerpunkte liegen auf dem Gebiet der digitalen Editorik und der digitalen Textanalyse, insbesondere im Bereich der mittelhoch-deutschen Literatur.

Thomas Weitin ist Professor für Germanistik und Digitale Literaturwissenschaft an der TU Darmstadt und leitet dort das *LitLab*. Zu seinen Forschungsschwerpunkten zählen die literaturwissenschaftliche Korpusanalyse sowie Mixed-Methods-Ansätze, die Statistik, Hermeneutik und kognitionswissenschaftliche Empirie verbinden.

Marcus Willand ist wissenschaftlicher Mitarbeiter an der Univ. Heidelberg, Redakteur der *Scientia Poetica* und Projektleiter im VW-geförderten *Mixed-Methods*-Projekt *QuaDrama*. Seine Forschungsinteressen sind Poetologien der dramatischen Figur, digitale Dramenanalyse und literaturwissenschaftliche Theorie- und Modellbildung.

Simone Winko ist Professorin für Neuere deutsche Literatur an der Universität Göttingen. Ihre Forschungsschwerpunkte liegen in der Literaturtheorie, den Themenfeldern ‚Literatur und Emotion‘, ‚Wertung und Kanonbildung‘ und ‚digitale Literatur‘.

Angelika Zirker ist akad. Rätin und apl-Professorin am Englischen Seminar der Universität Tübingen. Ihre Forschungsschwerpunkte liegen in der frühen Neuzeit sowie in der Literatur und Kultur des 19. Jahrhunderts. Sie hat TEASys (Tübingen Explanatory Annotations System) mitentwickelt und ist Betreuerin im Graduiertenkolleg 1808 Ambiguität: Produktion und Rezeption sowie PI im SFB 1391 Andere Ästhetik, u. a. im Projekt „Merkmale ästhetischer Reflexionsfiguren: Systematische Annotation und quantitative Analyse“.



Digitale Literaturwissenschaft. Zur Einführung

Fotis Jannidis

Der Begriff ‚digitale Literaturwissenschaft‘ ist, wie die vergleichbare Formulierung ‚digitale Geisteswissenschaften‘, eine Verlegenheitslösung, um allzu barocke Bezeichnungen zu vermeiden. Genauer wäre es, von der Literaturwissenschaft im Kontext der Digitalisierung und Vernetzung zu sprechen. Digitalisierung und Vernetzung betreffen den Gegenstand der Literaturwissenschaft, also literarische Texte im weitesten Sinne sowie die verschiedenen Formen gesellschaftlicher Kommunikation über diese Texte, ebenso wie die disziplinären Methoden der Aufbereitung des Gegenstands sowie der Analyse und die Schnittstellen des Fachs zu den angrenzenden Institutionen, insbesondere den Bibliotheken und Archiven. Nimmt man noch hinzu, dass es ‚die‘ Literaturwissenschaft nicht gibt, sondern vielmehr eine Vielzahl sehr unterschiedlicher Philologien mit sehr eigenen Traditionen – selbst innerhalb einer Nationalphilologie wie der germanistischen Literaturwissenschaft –, sodass selbst eine Formulierung wie ‚die germanistische Literaturwissenschaft‘ in erster Linie eine nützliche, aber eben auch stark homogenisierende Fiktion ist, dann wird klar, wie groß und divers das Feld ist, über das in diesem Band gesprochen werden soll. Man wird diese Diversität in einigen Jahren, wenn die Quellen entsprechend ausgewertet worden sind, vielleicht angemessener in ihren jeweiligen Häufigkeitsverteilungen beschreiben können. Die Aufgabe, die sich dieser Band gesetzt hat, ist demgegenüber altmodischer: Es geht darum, durch Reflexion auf den inzwischen erreichten jeweiligen Stand in ausgewählten Arbeitsfeldern Möglichkeiten und Probleme des Digitalen sichtbar zu machen. Die getroffene Auswahl privilegiert offensichtlich die Innovationen in unangemessener Weise, wenn unter ‚angemessen‘ verstanden wird, die Literaturwissenschaft insgesamt in Zeiten der Digitalisierung und Vernetzung in einer

F. Jannidis (✉)
Universität Würzburg, Würzburg, Deutschland
E-Mail: fotis.jannidis@uni-wuerzburg.de

Momentaufnahme zu beschreiben. Aber darum geht es uns nicht. Vielmehr soll das Momentum der Veränderung im Vordergrund stehen. Nicht weil dahinter die These steht, dass diese Veränderung bald das ganze Fach ergreifen wird, sondern weil es der Teil ist, der uns aufgrund seines Innovations- und Irritationspotentials besonders interessiert.

Bevor die Bereiche, in denen der vorliegende Band diese Veränderungen nach- und vorzeichnet, wenigstens ansatzweise skizziert werden, einige Worte zur Ausgangslage: Unter ‚Digitalisierung‘ oder ‚digitaler Transformation‘ versteht man den gesamten Prozess der gesellschaftlichen Veränderung bedingt durch die Verwendung von digitalen Informationen in einer vernetzten Welt. Auch wenn man erkannt hat, dass dieser Prozess nahezu alle Bereiche des Lebens erfasst, sind die Gestalten und Auswirkungen der Transformation nur in Ansätzen verstanden – und das betrifft auch die Wissenschaften. Wie so oft beim Verständnis von Neuem ist weder die Rhetorik der disruptiven Innovation noch die Gegenrhetorik, dass alles beim Alten bleibe, sehr hilfreich. Die neue Qualität dieser Welt beruht nicht auf einer ontologischen Differenz, sondern auf der enormen Beschleunigung der Prozesse. Digitale Informationen können, anders als analoge Informationen, in eine prinzipiell endlose Kette von sofortigen Verarbeitungsprozessen eingespeist werden. Man kann das Vorkommen eines Wortes in einer Millionen Seiten auch analog ermitteln oder die partiellen Ableitungen in einem neuronalen Netz auch mit der Hand berechnen, aber der Zeitaufwand wäre so groß, dass daraus keine neuen, allgemein verwendeten Praktiken entstehen könnten. Das gilt mutatis mutandis auch für die Vernetzung und die Geschwindigkeit des Informationstransports. Die Komplexität der Prozesse, die durch das Digitale und die Vernetzung möglich geworden sind, hat in den letzten 50 Jahren rasant zugenommen, sodass, wenn es auch an keiner Stelle einen Bruch mit der vorherigen Welt gibt, die Gegenwart paradoxerweise inzwischen doch fundamental anders aussieht.

Hier ist nun nicht der Ort, dem im Detail nachzugehen, aber es sei noch einmal festgehalten, dass an diesem Prozess von Anfang an auch Forschende aus den Geistes- und Kulturwissenschaften aktiv beteiligt waren,¹ auch wenn dies vielen Vertreterinnen und Vertretern dieser Fächer immer noch unbekannt ist und viele noch immer den Eindruck haben, dass die digitale Transformation ihre Fächer nicht wirklich betrifft. So stammen immerhin 0,4 % der Forschung zur Anwendung der künstlichen Intelligenz in den Jahren 1996–2016 aus den

¹Vgl. zur Geschichte der digitalen Geisteswissenschaften etwa die Überblicksdarstellungen von Hockey und Thaller: Susan Hockey, „The History of Humanities Computing“, in: Susan Schreibman/Ray Siemens/John Unsworth (Hg.), *A Companion to Digital Humanities*, o. O. 2007, 1–19. DOI: <https://doi.org/10.1002/9780470999875.ch1>; Manfred Thaller, „Geschichte der Digital Humanities“, in: Fotis Jannidis/Hubertus Kohle/Malte Rehbein (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart 2017, 3–12. DOI: https://doi.org/10.1007/978-3-476-05446-3_1. Siehe auch die Interviews, die Julianne Nyhan mit einigen Vertretern der ersten und zweiten Generation geführt hat: Julianne Nyhan/Andrew Flinn, *Computation and the Humanities. Towards an oral history of Digital Humanities*, Cham 2016.

Geisteswissenschaften.² Auch die Geschichte der digitalen Literaturwissenschaft beginnt bereits in den frühen 1960er Jahren, und sie hat seitdem einen festen Platz innerhalb der digitalen Geisteswissenschaft.³

Die Veränderungen der Literaturwissenschaften durch die Digitalisierung und Vernetzung betreffen vor allem vier Bereiche: den Gegenstand, die Erstellung und Verwendung von Editionen und anderen annotierten Textsammlungen, die quantitative Analyse digitaler Daten sowie die Kommunikation im Fach und mit Bibliotheken und Archiven. Diese vier Aspekte prägen auch die Grundstruktur dieses Bandes und jedem sind eine Einleitung und eine Reihe von Aufsätzen gewidmet. Daher sollen im Folgenden nur einige Stichworte soviel umreißen, dass sichtbar wird, wie umfassend die Veränderungen sind. Jeder dieser Bereiche ist heute auch von digitalen Arbeitsformen geprägt. Man mag diese Veränderungen für irrelevant halten und annehmen, sie würden nur das Äußerliche einer eigentlichen Arbeitsweise betreffen, aber das wäre medientheoretisch vielleicht etwas naiv. Denn auch wenn man nicht einer starken Variante der Medientheorie anhängt, die eine sehr enge Kopplung zwischen Medium und Inhalt annimmt, kann man hier vielfältige und komplexe Wechselwirkungen und Bedingtheiten beschreiben.

1. *Veränderungen des Gegenstandsbereichs.* Literarische Texte werden geschrieben, gedruckt, im Feuilleton rezensiert, und hin und wieder gibt es einen Preis – so gesehen scheint sich nichts geändert zu haben. Wenn man aber mit einem weiten Literaturbegriff auf die Gegenwartsgesellschaft blickt, mit einem Begriff, der alle Formen der fiktionalen ästhetischen Kommunikation umfasst, auch die verachteten und vernachlässigten Formen wie Internetmemes und Twittergeschichten, Gesangsvereinslieder und Fanfiction oder andere Formen nicht-professionellen Schreibens, dann wird das ganze Ausmaß der Veränderungen deutlich. Viele Menschen hören heute viel mehr Literatur als sie lesen, und sie verwenden dabei mit großer Selbstverständlichkeit digitale Vertriebsformen, die von der gelegentlichen Youtube-Lesung bis zu spezialisierten Anbietern wie Audible reichen.⁴ Ein substantieller Teil literarischer Produktion ist Teil fiktionaler Universen, die in verschiedenen Medien wie Kinofilmen, TV-Serien, Computerspielen, Hörspielen, Brettspielen

²Vgl. OECD, *The science behind AI. 1996–2016*, o. O. 2019. DOI: <https://doi.org/10.1787/57defaa7-en>.

³Im englischen Sprachraum wird das Feld meist als Digital Literary Studies bezeichnet; vgl. David L. Hoover/Jonathan Culpeper/Kieran O'Halloran (Hg.), *Digital Literary Studies. Corpus Approaches to Poetry, Prose, and Drama. The Corpus, the Computer and the Study of Literature*, New York 2008; Ray Siemens/Susan Schreibman (Hg.), *A Companion to Digital Literary Studies*, Malden/Oxford 2013.

⁴Vgl. dazu etwa Klaus Schenk/Ingold Zeisberger, *Literarisches Hören. Geschichte – Vermittlung – Praxis*, Kiel 2019.

usw. zugleich zugänglich sind.⁵ Hierzu ist wohl auch die überraschende Fülle an Fanfiction – Texte von Laienautoren, die in etablierten Universen angesiedelt sind – zu zählen, die im Internet zum einen erstmals breit wahrnehmbar wird, da sie früher nur in Ausnahmen gedruckt wurde, zugleich aber durch die vereinfachte Binnenkommunikation und größere Sichtbarkeit etwa in Foren enormen Zulauf erhält.⁶ Programmatische Formen des digitalen Schreibens etwa in Form von Hypertexten hatten in einem Avantgarde-Feld der Literatur einige Jahre Konjunktur, ohne dass sie – wie eigentlich alle Formen der Avantgarde – den Anspruch auf einen allgemeinen radikalen Wandel realisieren konnten.⁷ Aber auch der Schreibprozess selbst ist durch die Digitalisierung geprägt. Sei es bei einer sehr kleinen Zahl von Autorinnen und Autoren durch die radikale Verweigerung und das Bestehen auf Stift und Papier, sei es durch die Möglichkeiten des Digitalen, die schnelle Veränderbarkeit, die einfachere Wiederverwendbarkeit und die algorithmische Unterstützung. Das reicht von der Rechtschreibhilfe und dem Thesaurus bis zu Optionen, die sich vor allem schnell in den Bereichen des Schreibens etabliert haben, in denen dieses in komplexere Verwertungsketten eingebunden ist. Man denke etwa an die Fülle von Software zur Unterstützung beim Schreiben von Drehbüchern, die kollaboratives Schreiben, Materialorganisation, selektive Perspektiven auf das Geschriebene wie Figurenverzeichnisse u. a. m. ermöglicht.⁸ Ähnliches gilt etwa für das Schreiben von Texten für populäre Musik, die ja auch in sehr großer Zahl von Firmen und Fans gesammelt und publiziert werden.⁹ Und natürlich gehören hierher auch die immer leistungsfähigeren Programme, die die Rolle des professionellen Übersetzers immer stärker vom Autor zum Bearbeiter verschieben.

Die Kommunikation über Literatur hat sich grundlegend geändert, sei es die öffentlich sichtbare Kommunikation über Literatur in Online-Zeitschriften oder Blogs und Rezensionen bei *Amazon* oder etwa *Goodreads*¹⁰, sei es die interne zwischen Autoren und Lektoren und Agenten, seien es die Formen der Verlagswerbung und der Gestaltung von Informationsflüssen zwischen Verlagen und Publikum und auch die zahlreichen und viel genutzten neuen Formen der Selbstdarstellung von Autoren.¹¹ Die digitale Transformation beschleunigt hier oft

⁵Das Phänomen ist auch unter dem Stichwort ‚Transmediales Erzählen‘ bekannt; vgl. Claudia Söller-Eckert, „Transmediales Erzählen“, in: Matías Martínez (Hg.), *Erzählen. Ein interdisziplinäres Handbuch*, Stuttgart 2017, 108–110. DOI: https://doi.org/10.1007/978-3-476-05364-0_18.

⁶Vgl. z. B. Kristina Busse, *Framing Fan Fiction. Literary and Social Practices in Fan Fiction Communities*, Iowa City 2017.

⁷Vgl. etwa die Arbeiten von Roberto Simanowski; z. B. Roberto Simanowski, *Interfictions. Vom Schreiben im Netz*, Frankfurt a. M. 2002.

⁸Vgl. z. B. Screenwriting Software wie Final Draft oder Causality.

⁹Vgl. Lyric Writing Tools wie TuneSmith.

¹⁰Vgl. etwa die einschlägigen Beiträge in Andrea Bartl/Markus Behmer, *Die Rezension. Aktuelle Tendenzen der Literaturkritik*, Würzburg 2017.

¹¹Vgl. Elisabeth Sporer, *(Selbst-)Inszenierung von Autorinnen und Autoren im Internet am Beispiel von Autorenhomepages*, Baden-Baden 2019.

Prozesse, die schon deutlich länger laufen, unter diesen neuen Bedingungen aber eine zusätzliche Dynamik entfalten, etwa die Konzentrationsprozesse in der Verlagslandschaft, die durch die hohen Kosten der Umstellung auf eine durchgehend digitale Produktion beschleunigt werden.

Die digitale und vernetzte Kommunikation führt einerseits zu massiven Aufmerksamkeitskonzentrationen auf wenige Punkte, aber andererseits war es noch nie so leicht für die Liebhaberinnen und Liebhaber noch des entlegensten Gegenstands, Gesinnungsfreunde zu finden, wodurch eine sehr vielfältige und kaum überschaubare Kommunikationswelt entstanden ist. Literarisches partizipiert an beiden Tendenzen, sei es durch den schnell übersetzten Weltbestseller, sei es durch die Nischenpublikation oft im Eigenverlag auf digitalen Plattformen – mit den bekannten Ausnahmen wie etwa *Fifty Shades of Grey*, die vom letzteren zum ersten werden. Auch hier gilt, dass wir Vieles noch nicht wissen, wahrscheinlich Wesentliches noch nicht einmal sehen. Das gilt nicht nur für die Frage nach der Eigengesetzlichkeit der kleinen und kleinsten Kommunikationskreise, in denen sich ja u. a. auch die Liebhaber von schwierigen Autoren oder Gattungen wie experimenteller Lyrik bewegen, sondern auch für die Frage nach den Gesetzmäßigkeiten der neuen ‚Weltliteratur‘.¹²

So leicht es ist, auf all diese Veränderungen zu verweisen, so schwer, vielleicht unmöglich ist es, diese in einen überzeugenden und konsistenten Zusammenhang mit Formen und Inhalten von Literatur zu bringen. Dahinter verbirgt sich das klassische Problem jeder medien-, sozial- oder kulturgeschichtlichen Kontextualisierung von Literatur, für die es auch nach Kittler, Luhmann und Bourdieu keine integrative Theorielösung gibt, was ja nicht nur für die Gegenwart, sondern auch für alle anderen Epochen gilt. Der beobachtbaren Diversität am angemessensten erscheint aber wohl ein Modell, in dem Literatur und digitale Transformation in einem eher vermittelten Zusammenhang stehen, sodass man etwa im Fall von Fanfiction zwar die Existenz dieser Literatur vom Vorhandensein entsprechender digitaler Infrastrukturen abhängig machen kann, während die Inhalte und Formen dadurch aber nicht direkt bedingt sind, sondern eher von der sich so formierten Lese- und Schreibgemeinschaft.

2. *Edition und Annotation.* Historisch gesehen haben sowohl die Editionsphilologie als auch die quantitative Analyse erst durch die Digitalisierung und dann die Vernetzung eine gleichzeitige Dynamisierung erfahren, aber im Feld der Editionen ist dies zuerst sichtbar geworden.¹³ Bei digitalen Editionen wurde sehr

¹²Vgl. die Beiträge in dem Band David Damrosch (Hg.), *World Literature in Theory*, Chichester/Malden 2014.

¹³Das lässt sich schon am Publikationsdatum zentraler einschlägiger Veröffentlichungen ablesen: Wilhelm Ott, „Edition und Datenverarbeitung“, in: Herbert Kraft (Hg.), *Editionsphilologie*, Darmstadt 1990, 59–70. Peter Shillingsburg, *Scholarly editing in the computer age. Theory and practice*, Ann Arbor (MI) 1996; Lou Burnard/Katherine O’Brien O’Keeffe/John Unsworth (Hg.), *Electronic textual editing*, New York 2006; Patrick Sahle, *Digitale Editionsformen. 3 Bde.*, Norderstedt 2013 (= Schriften des Instituts für Dokumentologie und Editorik 7–9); Elena Pierazzo, *Digital Scholarly Editing. Theories, Models and Methods*, London/New York 2015.

früh mit automatischen Verfahren gearbeitet, erst vor allem bei der Erstellung der Editionen, z. B. bei der Kollationierung oder dem Management von Varianten, dann ab den 1990ern zunehmend auch bei der Publikation der Editionen, sodass heute die hybride Edition, also die Publikation gleichzeitig im Druck und digital, zum Modell geworden ist, da auf diese Weise der vielfältigen Nutzung einer Edition als Arbeitsmittel am besten gedient zu sein scheint. Durch das Vorliegen eines von Philologen entwickelten und auch gepflegten Standards zur Annotation von Editionen, den *Guidelines der Text Encoding Initiative*,¹⁴ hat dieses Feld nach Jahren beschleunigter Entwicklung eine gewisse Reife erreicht, was sich unter anderem daran zeigt, dass inzwischen eine ganze Reihe von digitalen Editionen abgeschlossen vorliegen.¹⁵ Neuere Forschungsfragen betreffen zumeist avancierte Aspekte digitaler Editionen, etwa die Analyse von Varianten mit quantitativen Verfahren oder deren alternative Modellierung in Form von Graphen, oder neue Probleme, z. B. wie man die oft sehr spezifischen Funktionalitäten solcher Editionen ebenso nachhaltig bewahren kann wie ihren digitalen Text.

Unter technischer Perspektive ist die Anreicherung eines Texts mit Informationen zu Varianten eine von vielen möglichen Formen der Annotation. Literaturwissenschaftliche Annotationen jenseits der Edition finden sich etwa im Bereich der Narratologie. Dabei zeigen sich bislang vor allem zwei Interessen: Zum einen geht es darum, literaturwissenschaftliche Begriffe auf den Prüfstand zu stellen, indem man sie nicht nur an einigen typischen Beispielen exemplifiziert, sondern auch erprobt, wie brauchbar sie sich in einer durchgängigen, genauen Textanalyse erweisen.¹⁶ Zum anderen ist die automatische Annotation, die für viele Fragen die Vorbedingung einer quantitativen Analyse ist, auf solche annotierten Korpora angewiesen, um entsprechende Werkzeuge mit Verfahren des maschinellen Lernens erstellen zu können.¹⁷ Hierbei ergibt sich mit der messbaren Übereinstimmung der Annotatoren ein interessanter Indikator für die Komplexität eines Begriffs, aber auch für den Grad an Operationalisierbarkeit.

3. *Quantitative Analyse*. Die Analyse literarischer Texte mit Verfahren der Statistik und des maschinellen Lernens ist wohl der Bereich, der sich in den letzten zehn

¹⁴TEI Consortium, eds. *TEI P5. Guidelines for Electronic Text Encoding and Interchange*. [Version 4.0.0]. [13.02.2020]. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/> (Stand: 26.08.2011).

¹⁵Vgl. etwa die Liste mit laufenden und abgeschlossenen Editionen unter <http://www.digitale-edition.de/> (Stand 01.08.2020).

¹⁶Vgl. z. B. Evelyn Gius/Janina Jacke, „Informatik und Hermeneutik. Zum Mehrwert interdisziplinärer Textanalyse“, in: *Sonderband der Zeitschrift für digitale Geisteswissenschaften* 1 (2015). DOI: https://doi.org/10.17175/sb001_006.

¹⁷Vgl. z. B. Markus Krug u. a., „Description of a Corpus of Character References in German Novels – DROC [Deutsches Roman Corpus]“, in: *DARIAH-DE Working Paper*, Göttingen 2018. Oder Annelen Brunner u. a., „Corpus REDEWIEDERGABE“, in: o.Hg., *Proceedings of The 12th Language Resources and Evaluation Conference*, Marseille (France) 2020, 803–812, <https://www.aclweb.org/anthology/2020.lrec-1.100>.

Jahren am schnellsten entwickelt hat und der auch besonders umstritten ist. Auch dieses Feld existiert schon seit den 1960er Jahren und kann auf Denk- und Methodentraditionen zurückgreifen, die bereits in der ersten Hälfte des 20. Jahrhunderts Verbreitung fanden. Das gilt insbesondere für die Autorschaftsattribuion, die schon vor der Ankunft des Computers an der numerischen Erhebung von Daten interessiert war und lange Zeit – neben der Erstellung von Konkordanzen – die wichtigste Anwendung des Computers war.¹⁸ Von den 1960er bis zu den 1990er Jahren konzentrierte sich die Analyse vor allem auf einen Text oder eine kleine Gruppe von Texten; neben der Autorschaftsattribuion¹⁹ beschäftigte man sich mit allgemeinen stilistischen Analysen, der Metrik, verwendeten Wortfeldern und dem Thema der Texte.²⁰

Die Beschränkung auf wenige Texte hatte pragmatische Gründe. ‚Rechenanlagen‘ waren noch keineswegs allgemein zugänglich, vor allem aber war die Digitalisierung der Texte mit erheblichem Zeitaufwand verbunden. Die Rezeption dieser Arbeiten war insgesamt wenig enthusiastisch, nicht zuletzt wohl, weil das notwendige Spezialwissen für den Umgang mit dem Computer und den mathematischen Werkzeugen in keinem sinnvollen Verhältnis zum Ertrag zu stehen schien, diese Analysen sich also vor allem für diejenigen (häufig Fachfremden) anboten, die ohnehin über diese Kompetenzen verfügten. Das änderte sich deutlich mit der allgemeinen Verbreitung von Personal Computern, mit der Verfügbarkeit großer Textsammlungen und dem Vorhandensein einer Tradition im Umgang mit digitalen literarischen Texten im Feld der digitalen Geisteswissenschaften. Zuerst wurde diese Richtung unter der Bezeichnung ‚Distant Reading‘²¹ weithin wahrgenommen und auch über den engeren Kreis der Beteiligten hinaus intensiv und kontrovers diskutiert.²² In den letzten Jahren werden unter der Bezeichnung *Computational Literary Studies* die verschiedenen Strömungen zusammengeführt.²³ Gemeinsam ist den meisten Arbeiten, dass sie größere

¹⁸Siehe Susan Hockey, *A guide to computer applications in the humanities*, Baltimore/London 1980, 122.

¹⁹Vgl. die Forschungsüberblicke in Patrick Juola, „Authorship Attribution“, in: *Foundations and Trends in Information Retrieval* 1/3 (2006), 233–334. Und: Efstathios Stamatatos, „A survey of modern authorship attribution methods“, in: *J. Am. Soc. Inf. Sci. Technol* 60/3 (2009), 538–556. DOI: <https://doi.org/10.1002/asi.v60:3>.

²⁰Vgl. etwa die einschlägigen Arbeiten in Helmut Kreuzer/Rul Gunzenhäuser (Hg.), *Mathematik und Dichtung*, München 1965.

²¹Vgl. Franco Moretti, *Graphs, Maps, Trees. Abstract Models for Literary History*, London/New York 2007; Franco Moretti, *Distant Reading*, London/New York 2013; Matthew L. Jockers, *Macroanalysis. Digital Methods and Literary History*, Urbana 2013.

²²Vgl. etwa den Angriff von Nan Z. Da und die anschließende Diskussion im *Journal of Cultural Analytics* und im *Critical Inquiry*; Nan Z. Da, „The Computational Case against Computational Literary Studies“, in: *Critical Inquiry* 45/3 (2019), 601–639. DOI: <https://doi.org/10.1086/702594>. Sowie: <https://culturalanalytics.org/section/1580-debates>.

²³Vgl. etwa das DFG Schwerpunktprogramm ‚Computational Literary Studies‘, https://www.dfg.de/foerderung/info_wissenschaft/2018/info_wissenschaft_18_30/.

und sehr große Textsammlungen verwenden und sich dabei moderner Verfahren des maschinellen Lernens, der Statistik und der natürlichen Sprachverarbeitung bedienen. Seit dem Anfang der 2010er Jahre wurden diese Verfahren durch *Deep Learning*, also den Einsatz komplexerer neuronaler Netze, und große vortrainierte Modelle auf der Basis von sehr vielen Daten deutlich verbessert, was einen weltweiten Forschungsboom ausgelöst hat, der auch die *Computational Literary Studies* dynamisiert hat.

Die Bandbreite der verschiedenen Methoden ist inzwischen ausgesprochen groß, wenn es auch immer noch zahlreiche Aspekte literarischer Texte gibt, die man nicht direkt erfassen kann, z. B. die Handlung eines Erzähltextes. Gemeinsam ist den Verfahren, dass sie für die Einzeltexte und die gesamte Textsammlung eine präzise Beschreibung unter genau einer Perspektive liefern. Das ist so ziemlich komplementär zur Erfassung eines Textes durch eine Leserin oder einen Leser, die im Prozess der Lektüre auf der Basis von Weltwissen einschließlich ihres Wissens über Literatur ein reiches Modell erstellen. Insofern ist es wenig überraschend, dass fast alle Forschenden in diesem Feld ihre Arbeiten nicht als ein Ersetzen der alten Literaturwissenschaft durch eine neue projektieren, sondern sie vielmehr als eine Erweiterung des literaturwissenschaftlichen Methodenvorrats ansehen. Der große Vorteil dieser Methoden besteht eben vor allem darin, dass sehr große Textsammlungen in den Blick genommen werden können, wenn auch unter einer sehr reduzierten Perspektive, und dass dadurch etwa Phänomene sichtbar werden, die aufgrund ihrer zeitlichen Ausdehnung bislang nicht wahrgenommen wurden²⁴ oder die die große Menge von Texten betreffen, über die zumindest das Fach wenig oder nichts weiß, z. B. Heftrömene, Fanfiction oder die Novellenliteratur des 19. Jahrhunderts jenseits des Höhenkamms.

Keineswegs identisch mit diesem Aspekt ist ein anderer, auch wenn es im Personal und den Arbeitsweisen große Überschneidungen gibt: die Verwendung von Denkmodellen der empirischen Wissenschaften. Darunter verstehe ich hier stark vereinfachend die Ausrichtung an einem Denk- und Arbeitsmodell, dessen Kern die Falsifikation ist, vor allem die inferenzstatistischen Verfahren zur Prüfung von Hypothesen. Explorative Verfahren haben in diesem Kontext einen genau definierten Ort, indem sie zur Generierung von Hypothesen eingesetzt werden, aber selbst noch kein gesichertes Wissen produzieren, sondern nur eine Vorstufe in einem Prozess darstellen, der letztlich in eine Hypothesenprüfung einmündet. Empirisch in diesem Sinne sind bislang nur wenige Untersuchungen in der quantitativen Literaturwissenschaft, vielmehr kommen dort fast durchgehend explorative Verfahren zur Anwendung, deren Ergebnisse dann eher qualitativ unter Einbeziehung von Fachwissen mit möglichen Erklärungen verbunden werden. Das ist der Skepsis einiger Vertreterinnen und Vertreter gegenüber empirischen Wissenschaftsmodellen geschuldet, aber auch dem Zustand der diversen Textsammlungen, die eben nicht randomisiert aus einer Gesamtpopulation

²⁴Vgl. Ted Underwood, *Distant Horizons*, Chicago 2019.

zusammengestellt wurden, sondern durch komplexe Auswahlprozesse geprägt sind, in die wiederum Bibliothekserwerbungsstrategien und Selektionskriterien für Digitalisierungskampagnen eingehen und einen undurchschaubaren Bias erzeugen. Zugleich aber folgt die Anwendung quantitativer Verfahren aufgrund ihres Entstehungskontexts fast automatisch einer Drift in Richtung empirischer Wissenschaftsmodelle. Die Literaturwissenschaften gehören – zusammen mit den Musik- und Kunstwissenschaften – zu den letzten Fächern, in denen solche Modelle Fuß fassen, und ein Blick auf die Geschichte von Fächern wie Psychologie, Soziologie oder Linguistik zeigt, wie unterschiedlich die Adaptionen verlaufen können, dass sie aber höchst spannungsreich waren und oft noch sind. In den Sozialwissenschaften wird die Methodenlandschaft grundlegend in qualitative und quantitative Verfahren geordnet, wobei dort unterstellt wird, dass ‚quantitativ‘ und ‚empirisch‘ gleichbedeutend sind. In letzter Zeit mehren sich die Stimmen in den Sozialwissenschaften, die einen pragmatischen Umgang mit diesen Methodenangeboten im Kontext von Mixed-Methods-Designs fordern.²⁵ Ob die Diskussion der quantitativen Verfahren in den Literaturwissenschaften von solchen Perspektiven profitieren kann, ist zur Zeit noch unklar.

4. *Schnittstellen.* Die digitalen Gegenstände und Verfahren verändern die Ansprüche an die Einrichtungen, mit denen die Literaturwissenschaften kooperieren. Die Bibliotheken sind etwa durch die Digitalisierung schon lange in einem tiefgreifenden Umbauprozess begriffen,²⁶ dessen genauer Umfang noch nicht ganz abzusehen ist, da die Arbeitsteilung zwischen Rechenzentren und Bibliotheken häufig noch weniger von den Institutionen als von den beteiligten Persönlichkeiten und deren Zugang zu wesentlichen Ressourcen bestimmt ist. Aber es ist jetzt schon deutlich, dass Bibliotheken eine wesentliche Rolle bei der Herstellung, Pflege, Verwaltung und Erschließung von digitalen Informationen haben werden – und das betrifft auch die literaturwissenschaftlichen Editionen, Textsammlungen und die oben erwähnten neuen Forschungsdaten, die etwa bei der Erstellung digitaler Editionen anfallen. Ebenso klar ist, dass Bibliotheken ganz neue Informationsressourcen anbieten werden, z. B. Word Embeddings trainiert auf ihren großen Datenbeständen. Weniger klar jedoch ist, ob dies alle Bibliotheken gleichmäßig betrifft und wie sich die künftige Arbeitsteilung zwischen den Bibliotheken gestalten wird. Und auch die Frage, wie mit etwas ferner liegenden Aspekten der digitalen Methodenerweiterung umzugehen ist, ist noch ungelöst, etwa wo die Forschungssoftware, die bei der Erstellung von Editionen oder von quantitativen Untersuchungen verwendet wird, nachhaltig vorgehalten wird. Nicht zuletzt haben sich manche Bibliotheken als wichtige Agenten bei dem Versuch etabliert, die Wissenschaftskommunikation trotz der

²⁵Vgl. z. B. Udo Kelle, „Mixed Methods“, in: Nina Baur/Jörg Blasius (Hg.), *Handbuch Methoden der empirischen Sozialforschung*, Berlin/New York 2014, 153–166.

²⁶Vgl. z. B. Heike Neuroth u. a. (Hg.), *Evolution der Informationsinfrastruktur. Kooperation zwischen Bibliothek und Wissenschaft*, Glückstadt 2013.

massiven Konzentrationsprozesse bei den Verlagen durch Open Access und andere Strategien zugänglich zu halten.

Gerade im Bereich der Editionen hat sich inzwischen die intensivste Vernetzung mit Institutionen des kulturellen Gedächtnisses ergeben. Das betrifft zum einen den automatischen Rückverweis der Editionen auf die verwendeten Quellen, die etwa in Archiven und Bibliotheken liegen, zum anderen aber auch die Integration der Editionen und der Teilergebnisse in übergeordnete Ressourcen, etwa wenn Editionen angereicherte Handschriftenbeschreibungen erstellen, die auch gut in einem allgemeinen Handschriftenkatalog aufgehoben wären, oder wenn in deren Kontext bisher unbekannte Personen identifiziert und beschrieben werden, die auch in so etwas wie die Gemeinsame Normdatei der Deutschen Nationalbibliothek passen würden. Und nicht zuletzt stellt sich natürlich die Frage, wo die Editionen selbst langfristig liegen sollen. Das ist eine Frage, deren Antwort wohl erst mit dem Aufbau nationaler Infrastrukturen für Forschungsdaten gefunden sein wird.

Die digitale Transformation hat, wie jede Veränderung, viel Kritik auf den Plan gerufen. Da wird etwa die Angst geäußert, dass aus Computerspielern Massenmörder werden, oder das Internet zum endgültigen Verfall der Bildung führt. Die Kritik an der digitalen Literaturwissenschaft folgt ebenfalls einer kleinen Reihe von Mustern, die immer wieder aufleben. Ein erheblicher Teil der kritischen Stimmen unterstellt dabei dem Neuen, dass es das Alte ersetzen will. So war die Rezeption der ersten digitalen Editionen noch von der entsetzten Frage gefärbt, ob denn nun das Buch verschwinden würde, auch wenn die Befürworter digitaler Edition fast durchgängig die funktionale Differenzierung der Medien beschworen haben, die sich bislang auch weitgehend etabliert hat. Kritiker der quantitativen Literaturwissenschaft unterstellen beharrlich, sie wolle die herkömmlichen Methoden ersetzen, auch wenn praktisch in jedem längeren Text zu diesem Thema vonseiten der Befürworter betont wird, dass es sich um komplementäre Ansätze handle, die den Methodenvorrat erweitern würde – und das beginnt in Deutschland schon 1965 mit Kreuzers Einleitung zum Band *Mathematik und Dichtung*.²⁷ Im Folgenden sollen einige häufig wiederholte Argumente der Kritik aufgegriffen werden. Da es hier aber weniger auf die einzelne Stimme ankommt, sondern vielmehr auf das typische Argument, wird auf einen Einzelnachweis verzichtet.

„Das wussten wir bereits“. Das Ergebnis einer quantitativen Untersuchung sei deshalb irrelevant, weil das Fach das schon ‚wisse‘. Offensichtlich hängt hier alles vom Begriff des Wissens ab. Wenn man jede Proposition in Bezug auf einen literarischen Text bereits als Wissen bezeichnet, dann ist es wohl sehr schwierig, Neues zu sagen, insbesondere angesichts der zahlreichen Interpretationen, die kanonischen Texten gewidmet sind. Allerdings würde man damit diese Irrelevanz auch einem großen Teil der publizierten Interpretationen zuschreiben, da auch diese in vielen Fällen keine neuen Thesen aufstellen, sondern alte Thesen

²⁷ Helmut Kreuzer, „Mathematik und Dichtung. Zur Einführung“, in: Helmut Kreuzer/Rul Gunzenhäuser (Hg.), *Mathematik und Dichtung*, München 1965, 9–20, hier 17.

aufgreifen und z. B. durch neue Argumente zu stützen suchen. Kennzeichnend für die literaturwissenschaftliche Forschung ist doch gerade die Fülle von Interpretationsthesen insbesondere zu den kanonischen Texten, eine Fülle von Aussagen, die sich zum gehörigen Teil widersprechen. Was bedeutet es also, wenn in diesem Kontext durch quantitative Forschung eine solche These noch einmal aufgegriffen und mit quantitativen Argumenten begründet wird? Auf diese Weise wird die bekannte These neu gestützt und zwar aus einer methodisch andersartigen Perspektive. In den Sozialwissenschaften wird dieses Phänomen im Kontext der Diskussion um ‚Mixed Methods‘ als ‚Triangulation‘ bezeichnet und von vielen Vertretern eines pragmatischen Umgangs mit qualitativen und quantitativen Verfahren als Erkenntnisgewinn gewertet.²⁸

Ein weiterer Vorwurf lautet: ‚Die Themen sind veraltet‘. Das bezieht sich darauf, dass die Forschung mit quantitativen Verfahren sich tatsächlich auffallend häufig mit Begriffen wie ‚Autor‘ oder ‚Gattung‘ oder ‚Literaturgeschichte‘ beschäftigt, die ein anderer Teil der Literaturwissenschaft schon offiziell ad acta gelegt zu haben scheint (auch wenn sich diese theoretische Verabschiedung kaum in der Alltagspraxis niederschlägt).²⁹ Aber diese Verabschiedungen sind auch innerhalb der nicht-quantitativ arbeitenden Literaturwissenschaft keineswegs allgemein akzeptiert. Vielmehr ist es typisch für die pluralistischen Literaturwissenschaften der letzten 30 Jahre, dass es ein Nebeneinander von ganz diversen und auch widersprüchlichen Positionen gibt, sodass neben der Verabschiedung des Autorkonzepts in Nachfolge von Foucault eine florierende Autorenphilologie mit Handbüchern und runden Geburtstagen existiert. Aber selbst wenn man dies unberücksichtigt lässt, ergibt sich bei genauerer Betrachtung ein sehr viel komplexeres Bild. Vertreterinnen und Vertreter der quantitativen Verfahren verwenden die umstrittenen Begriffe ja nicht einfach, vielmehr setzen sie sich immer wieder kritisch mit den Positionen auseinander, die jene für obsolet erklären. Das sichtbarste Beispiel dafür ist wahrscheinlich die Stilometrie, die schon recht früh Einspruch gegen Foucaults Thesen zur Autorschaft erhoben hat.³⁰ Rezente Hinweise darauf, dass quantitative Verfahren gerade dafür geeignet sind, sehr langfristige Trends sichtbar zu machen, die der Literaturwissenschaft so bislang entgangen sind, lassen sich als selbstbewusste Kritik an einer Verabschiedung von Literaturgeschichte auffassen.³¹ Diese Revisionen von Positionen, die Teilen der

²⁸Vgl. Kelle (Anm. 25).

²⁹Vgl. Simone Winko, „Autor-Funktionen. Zur argumentativen Verwendung von Autorkonzepten in der gegenwärtigen literaturwissenschaftlichen Interpretationspraxis“, in: o.Hg., *Autorschaft. Positionen und Revisionen. Akten des DFG-Symposiums Salza*, Stuttgart 2002, 334–354.

³⁰Vgl. etwa schon die Arbeiten von Burrows sowie Craig und Kinney; John Burrows, „Computers and the Idea of Authorship“, in: Fotis Jannidis/Gerhard Lauer/Matias Martinez/Simone Winko (Hg.), *Rückkehr des Autors. Zur Erneuerung eines umstrittenen Begriffs*, Tübingen 1999, 133–144; Hugh Craig, „Style, statistics, and new models of authorship“, in: *Early Modern Literary Studies* 15/1 (2009), 1–42.

³¹Vgl. Underwood (Anm. 24).

Literaturwissenschaft lieb geworden sind, wird man wohl nicht mit dem Hinweis abweisen können, dass man doch schon seit diesem oder jenem Meisterdenker wisse, dass ein bestimmter Begriff nicht haltbar sei. Allerdings ist die Kritik an diesen etablierten Denkmodellen der Literaturwissenschaft durch Vertreterinnen und Vertreter quantitativer Verfahren bislang ein einseitiges Gespräch.

‚Quantitative Verfahren können das Wesentliche an Literatur niemals erfassen‘. Diese Meinung wird von Vielen geteilt, selbst wenn hier noch nichts über die jeweiligen Gründe ausgesagt wird, also was das Wesentliche von Literatur ist und warum sich das einer quantitativen Erfassung entzieht. In den verschiedenen Anwendungen des Arguments lassen sich einige typische Begründungen finden, z. B. weil literarische Texte zu komplex sind, weil sie immer mehrdeutig sind oder weil sie einzigartige Individualitäten sind. Andere Begründungen beziehen sich darauf, dass die Aktualisierung eines Textes und seine angemessene Bewertung bzw. eine ästhetische Wertschätzung nur von Menschen vorgenommen werden kann. In den jeweiligen Begründungen zeigt sich zumeist, dass der hier verwendete Literaturbegriff hoch evaluativ ist und nicht deskriptiv: Es geht nicht um alle fiktionalen, ästhetisch geformten Texte vom Hefroman über das politische Gelegenheitsgedicht, die Kabarettnummer bis zur künstlerischen Höchstleistung, für die häufig Werke wie Kafkas Roman *Der Prozeß* oder James Joyce *Ulysses* angeführt werden, sondern nur um diese letzte Gruppe. Den meisten Arbeiten im Bereich der quantitativen Analyse von Literatur liegt dagegen ein weiter, deskriptiver Literaturbegriff zugrunde. Wertung wird in diesem Kontext nur zum Thema, wenn Wertungen von Lesern mit Eigenschaften der Texte korreliert werden sollen,³² aber nicht als Leistung der verwendeten Verfahren. Die Perspektive auf den Einzeltext ist, methodisch bedingt, höchst selektiv, aber diese Einzelperspektive kann eben zugleich auf sehr viele Texte gerichtet werden. Dadurch entstehen Aussagen und Einsichten, die bislang in der Form kaum möglich waren und die den Methodenvorrat der an Methoden ja schon reichen Literaturwissenschaft noch einmal deutlich vergrößern. Daraus kann man wohl weder eine Heils- noch eine Untergangsgeschichte ableiten, aber es fällt leicht zu versprechen, dass viele interessante Ergebnisse noch kommen werden, sind doch auf diesem Weg die ersten Schritte kaum getan. Einen wichtigen Beitrag dazu kann auch die Modellierung leisten, die zum Kern der digitalen Literaturwissenschaft gehört.

Die vielfältigen Phänomene, die mit dem Begriff ‚digitale Literaturwissenschaft‘ in den Blick genommen werden, haben in vielen Aspekten mehr mit ihren analogen Gegenständen gemein als mit anderen digitalen. Die digitale Editionsphilologie etwa hat in vielerlei Hinsicht mehr Gemeinsamkeiten mit der traditionellen Editionsphilologie als mit den Veränderungen des literarischen Markts durch die digitale Transformation. Insofern suggeriert dieser Begriff eine Einheit, die man in der Sache so umfassend nicht finden wird. Zugleich aber

³²Vgl. den Beitrag von van Dalen in diesem Band.

gibt es eine Reihe von Aspekten, die tatsächlich alle diese digitalen Arbeitsfelder verbindet. Dazu gehört offensichtlich die gemeinsame Bedingtheit durch die Digitalisierung und die Vernetzung. Aber das betrifft auch und zentral den Aspekt der Modellierung. Mit ‚Modellierung‘ ist hier eine formale Beschreibung gemeint, die von der einfachen Datenbank eines Projekts, das Veränderungen im Gegenstandsbereich erfassen möchte, über die Abbildung von Texten im Workflow einer digitalen Bibliothek oder die fein gegliederte Modellierung einer Edition, die sich etwa in der XML/TEI-Kodierung niederschlägt, bis hin zu komplexen mathematischen Modellierungen etwa bei der quantitativen Analyse reicht. Stets muss der Gegenstand auf ein System klarer, letztlich auch maschinenlesbarer Kategorien abgebildet werden. Diese Tätigkeit ist – im Kontext der digitalen Geisteswissenschaften – als eine wesentliche Leistung des Arbeitsfeldes erkannt worden,³³ und natürlich ist gerade die Modellierung von Unschärfe und Vagheit schon länger Gegenstand einschlägiger Diskussionen, da nicht nur die Literaturwissenschaften diese für typische Eigenschaften ihrer Gegenstände erachten.³⁴ Unter dieser Perspektive gibt es also eine Einheit in der digitalen Literaturwissenschaft. Die doppelte Struktur des Feldes, einmal als zerstreute Komplemente der analogen Teilfelder und dann als durch die Techniken, die Praktiken und die notwendigen Wissensvoraussetzungen erzeugte Einheit, prägt es und trägt angesichts der enormen Geschwindigkeit des Wandels in allen Bereichen zu seiner schnell sich steigernden Komplexität bei. Die formale Modellierung erlaubt eine neue Beschreibung für literarische und literaturwissenschaftliche Phänomene, und die Literaturwissenschaft als soziales System erhält dadurch eine reichere Semantik.

Der Band versammelt die Beiträge des DFG-Symposiums ‚Digitale Literaturwissenschaft‘, das vom 9.–13. Oktober 2017 in der Villa Vigoni stattfand. Die Sektionsverantwortlichen Jan Christoph Meister, Andrea Rapp, Thomas Stäcker und Simone Winko haben bei der Auswahl der Beiträge und der Gestaltung der Diskussion ganz wesentlich mitgewirkt. Dem Format der germanistischen Symposien treu bleibend werden auch hier die Diskussionen wiedergegeben, was nur durch das mühsame Geschäft der Protokollantinnen und Protokollanten Constanze Baum, Luise Borek, Jan Borkowski und Jan Horstmann möglich war. Carolin Hahn hat mit den Autoren und der Herstellung des Verlags kommuniziert und mit viel Sinn fürs Detail den Band für den Druck eingerichtet; letzte Arbeiten zur Fertigstellung des Bandes hat Agnes Hilger übernommen. Das Symposium

³³Ausgehend von Überlegungen von Willard McCarty, einem der Pioniere der Digital Humanities, ist dies inzwischen zu einem Leitthema der digitalen Geisteswissenschaften geworden; vgl. z. B. Willard McCarty, *Humanities Computing*, Basingstoke (Hampshire) 2005; Julia Flanders/Fotis Jannidis, *The Shape of Data in Digital Humanities. Modeling Texts and Text-Based Resources*, o. O. 2018 (= Digital Research in the Arts and Humanities), <https://books.google.de/books?id=XJF1jwEACAAJ>; Ariana Ciula u. a. (Hg.), *Models and Modelling between Digital and Humanities – A Multidisciplinary Perspective*, Cologne 2018 (= Historical Social Research 31).

³⁴Vgl. z. B. Michael Piotrowski, „Accepting and Modeling Uncertainty“, in: *Zeitschrift für digitale Geisteswissenschaften* (2019).

ist im Rahmen der Reihe der germanistischen Symposien von der DFG gefördert worden. Von Seiten des Verlags haben Ute Hechtfisher und Ferdinand Pöhlmann mit viel Geduld das Zustandekommen des Bandes unterstützt und begleitet. Allen sei hier ganz ausdrücklich für ihre vielfältigen Formen der Mitarbeit und Mitgestaltung gedankt. Für die Fehler und dafür, dass der Band sehr viel später als erhofft erscheint, ist eine unglückliche Verkettung von persönlichen und weltweiten Ausnahmesituationen verantwortlich. Die Autorinnen und Autoren haben die Möglichkeit erhalten, ihre Beiträge durch eine Ergänzung zu aktualisieren.³⁵

Literatur

- Bartl, Andrea/Behmer, Markus, *Die Rezension. Aktuelle Tendenzen der Literaturkritik*, Würzburg 2017.
- Burnard, Lou/O'Brien O'Keeffe, Katherine/Unsworth, John (Hg.), *Electronic textual editing*, New York 2006.
- Burrows, John, "Computers and the Idea of Authorship", in: Gerhard Lauer/Fotis Jannidis (Hg.), *Rückkehr des Autors. Zur Erneuerung eines umstrittenen Begriffs*, Tübingen 1999, 133–144.
- Busse, Kristina, *Framing Fan Fiction. Literary and Social Practices in Fan Fiction Communities*, Iowa City 2017.
- Ciula, Ariana u. a. (Hg.), *Models and Modelling between Digital and Humanities. A Multidisciplinary Perspective*, Cologne 2018.
- Craig, Hugh, "Style, statistics, and new models of authorship", in: *Early Modern Literary Studies* 15/1 (2009), 1–42.
- Da, Nan Z., "The Computational Case against Computational Literary Studies", in: *Critical Inquiry* 45/3 (2019), 601–639.
- Damrosch, David (Hg.), *World Literature in Theory*, Chichester 2014.
- Flanders, Julia/Jannidis, Fotis, *The Shape of Data in Digital Humanities. Modeling Texts and Text-Based Resources*, o.O. 2018 (= Digital Research in the Arts and Humanities), <https://books.google.de/books?id=XJF1jwEACAAJ>.
- Gius, Evelyn/Jacke, Janina, „Informatik und Hermeneutik. Zum Mehrwert interdisziplinärer Textanalyse“, in: *Sonderband der Zeitschrift für digitale Geisteswissenschaften* 1 (2015). DOI: https://doi.org/10.17175/sb001_006.
- Hockey, Susan, *A guide to computer applications in the humanities*, Baltimore/London 1980.
- Hockey, Susan, "The History of Humanities Computing", in: Susan Schreibman/Ray Siemens/John Unsworth (Hg.), *A Companion to Digital Humanities*, o.O. 2007, 1–19. DOI: <https://doi.org/10.1002/9780470999875.ch1>.
- Hoover, David L./Culpeper, Jonathan/O'Halloran, Kieran (Hg.), *Digital Literary Studies. Corpus Approaches to Poetry, Prose, and Drama. The Corpus, the Computer and the Study of Literature*, New York 2008.
- Jockers, Matthew L., *Macroanalysis. Digital Methods and Literary History*, Urbana 2013.
- Juola, Patrick, "Authorship Attribution", in: *Foundations and Trends in Information Retrieval* 1/3 (2006), 233–334.
- Kelle, Udo, "Mixed Methods", in: Nina Baur/Jörg Blasius (Hg.), *Handbuch Methoden der empirischen Sozialforschung*, Berlin/New York 2014, 153–166.
- Kreuzer, Helmut/Gunzenhäuser, Rul (Hg.), *Mathematik und Dichtung*, München 1965.

³⁵Die Beiträge lagen mit ganz wenigen Ausnahmen Ende 2019 in der vorliegenden Form vor. Ein Beitrag wurde vom Autor zurückgezogen.

- Kreuzer, Helmut, „Mathematik und Dichtung. Zur Einführung“, in: Helmut Kreuzer/Rul Gunzenhäuser (Hg.), *Mathematik und Dichtung*, München 1965, 9–20.
- Krug, Markus u. a., “Description of a Corpus of Character References in German Novels - DROC [Deutsches ROman Corpus]”, in: *DARIAH-DE Working Papers*, Göttingen 2018.
- McCarty, Willard, *Humanities Computing*, Basingstoke (Hampshire) 2005.
- Moretti, Franco, *Graphs, Maps, Trees. Abstract Models for Literary History*, London/New York 2007.
- Moretti, Franco, *Distant Reading*, London/New York 2013.
- Neuroth, Heike u. a. (Hg.), *Evolution der Informationsinfrastruktur. Kooperation zwischen Bibliothek und Wissenschaft [erschienen im Rahmen des zehnjährigen Jubiläums der Abteilung Forschung & Entwicklung 2012]*, Glückstadt 2013.
- Nyhan, Julianne/Flinn, Andrew, *Computation and the Humanities. Towards an oral history of Digital Humanities*, Cham 2016.
- Ott, Wilhelm, „Edition und Datenverarbeitung“, in: Herbert Kraft (Hg.), *Editionsphilologie*, Darmstadt 1990, 59–70.
- Pierazzo, Elena, *Digital Scholarly Editing. Theories, Models and Methods*, London/New York 2015.
- Sahle, Patrick, *Digitale Editionsformen. 3 Bde.*, Norderstedt 2013.
- Schenk, Klaus/Zeisberger, Ingold, *Literarisches Hören. Geschichte – Vermittlung – Praxis*, Kiel 2019.
- Shillingsburg, Peter, *Scholarly editing in the computer age. Theory and practice*, Ann Arbor (MI) 1996.
- Siemens, Ray/Schreibman, Susan (Hg.), *A Companion to Digital Literary Studies*, Malden/Oxford 2013.
- Simanowski, Roberto *Interfictions. Vom Schreiben im Netz*, Frankfurt a. M. 2002.
- Söllner-Eckert, Claudia, „Transmediales Erzählen“, in: Matías Martínez (Hg.), *Erzählen. Ein interdisziplinäres Handbuch*, Stuttgart 2017, 108–110.
- Sporer, Elisabeth, *(Selbst-)Inszenierung von Autorinnen und Autoren im Internet am Beispiel von Autorenhompages*, Baden-Baden 2019.
- Stamatatos, Efstathios, “A survey of modern authorship attribution methods”, in: *Journal of American Society for Information Science and Technology* 60/3 (2009), 538–556.
- Thaller, Manfred, „Geschichte der Digital Humanities“, in: Fotis Jannidis/Hubertus Kohle/Malte Rehbein (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart 2017, 3–12.
- Underwood, Ted, *Distant Horizons*, Chicago 2019.
- Winko, Simone, „Autor-Funktionen. Zur argumentativen Verwendung von Autorkonzepten in der gegenwärtigen literaturwissenschaftlichen Interpretationspraxis“, in: Heinrich Detering (Hg.), *Autorschaft. Positionen und Revisionen. Akten des DFG-Symposiums Salzwedel*, Stuttgart 2002, 334–354.

Online-Ressourcen

- Brunner, Annelen u. a., „Corpus REDEWIEDERGABE“, in: o.Hg., *Proceedings of The 12th Language Resources and Evaluation Conference*, Marseille (France), European Language Resources Association 2020, 803–812, <https://www.aclweb.org/anthology/2020.lrec-1.100>.
- OECD, *The science behind AI, 1996–2016*, o.O. 2019. DOI: <https://doi.org/10.1787/57defaa7-en>.
- Piotrowski, Michael: “Accepting and Modeling Uncertainty”, in: *Zeitschrift für digitale Geisteswissenschaften* 2019. DOI: https://doi.org/10.17175/sb004_006a.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.



Sektion I. Literatur und Literaturwissen- schaft unter digitalen Bedingungen



Einführung

Simone Winko

Die Sektion I versammelt Beiträge unter zwei unterschiedlichen Leitfragen. Zum einen konzentriert sie sich auf den Gegenstandsbereich der Literaturwissenschaft, auf *Literatur unter digitalen Bedingungen*. Wie hat sich das Phänomen Literatur seit der Verbreitung digitaler Medien gewandelt und, dies interessiert hier vor allem, welche Folgen ergeben sich aus den Veränderungen für die literaturwissenschaftliche Erforschung von Literatur? Zum anderen untersucht sie Konsequenzen für die Disziplin als solche und fragt nach der *Literaturwissenschaft unter digitalen Bedingungen*. Anders als in den folgenden Sektionen, die genauer begrenzte Aspekte der Veränderung in den Blick nehmen – editorische Praktiken, Methoden der Analyse von Literatur und die Schnittstellen –, geht es in dieser zweiten Gruppe um eine globalere Perspektive auf das Fach. Was ändert sich im Zeichen der Digitalität? Was kann oder was sollte es heißen, eine ‚digitale Literaturwissenschaft‘ zu betreiben und wo liegen die besonderen Herausforderungen?

(1) *Literatur unter digitalen Bedingungen*. Literarische Phänomene haben sich seit den 1980er Jahren langsam, aber kontinuierlich verändert, genauer gesagt: modifiziert und erweitert. Die Veränderungen kommen vor allem dann in den Blick, wenn man sich weder auf ein enges Konzept von Höhenkamm-literatur noch auf das einzelne Werk beschränkt und sich mit ästhetischen Abwertungen zurückhält. Passend für dieses weite Verständnis des Gegenstandes ist ein medienbezogener, auch Handlungen einschließender Literaturbegriff. ‚Literatur‘ bezeichnet dann ein Ensemble unterschiedlicher medialer Formate und Kommunikationsformen, eine vielfältige, dynamische Praktik, die weit mehr umfasst als gedruckte Einzelwerke und vom Feuilleton wahrgenommene

S. Winko (✉)
Universität Göttingen, Göttingen, Deutschland
E-Mail: simone.winko@phil.uni-goettingen.de

Autorinnen und Autoren und die Teil sowohl der aktuellen Partizipationskultur als auch von Medienverbänden sein kann. Viele der neuen Formate und Kommunikationsformen hängen mehr oder minder direkt mit dem digitalen Medienumbruch zusammen.

Zu diesem Fragenkomplex unseres Calls for Papers haben wir nur wenige passende Vorschläge bekommen. Vielleicht interessiert er Literaturwissenschaftlerinnen und Literaturwissenschaftler weniger oder wird nicht mit dem Schlagwort ‚digitale Literaturwissenschaft‘ verbunden. Da zudem noch Beiträge ausgefallen sind, sei im Folgenden ein Bild dessen skizziert, was die Veranstalterinnen und Veranstalter des Symposiums im Blick gehabt haben: Welche Phänomene könnten für die Frage nach einem erweiterten literaturwissenschaftlichen Gegenstandskonzept und Methodenspektrum berücksichtigt werden? Der Überblick kann hier nur kurz und ausschnitthaft ausfallen,¹ und auch die vermuteten Auswirkungen für die Literaturwissenschaft müssen im spekulativen Modus bleiben, da belastbare Aussagen dazu genaue Untersuchungen erfordern würden.

Der Vollständigkeit halber zu nennen sind die digitale Literatur, die in Form von Hyperfictions seit den 1980er Jahren erscheint und von experimentellen Texten bis zu solchen, die traditionelle Gattungsvorstellungen bedienen, ein weites Spektrum an Formen umfasst, sowie die „Computer-“ oder „Maschinenpoesie“, die literarische Texte durch Algorithmen erzeugt und ‚dichtet‘, indem sie auf verschiedene Weise importierten Inhalt nutzt: Sie kann die vom Nutzer eingegebenen Wörter mit internen strukturierten Wortlisten verbinden (wie z. B. der Gedicht-Generator *Poetron*) oder mit den Wörtern eines literarischen Werks arbeiten, die in neue syntaktische Zusammenhänge gebracht werden, etwa in Simon Biggs’ *The Great Wall of China* (1996), der Kafkas Nachlassfragment *Beim Bau der Chinesischen Mauer* verwendet. Sie erzeugen so mehr oder weniger sinnvolle, in jedem Fall aber ‚deutungsoffene‘ Texte von unterschiedlich hohem ästhetischen Anspruch. Auch textzentrierte und narrativ organisierte Adventuregames sind hier zu nennen.

Weniger diese ‚born-digital‘ literarischen Formen sind es aber, die die Literaturwissenschaft vor Herausforderungen stellen, als eher die vielen Spielarten der Netzliteratur. Ihr wesentliches Merkmal liegt darin, dass sie des WWW bedürfen, um produziert und rezipiert zu werden. Auch hier ist das Spektrum vorliegender Texte breit und reicht von Produkten, die das Netz in einem pragmatischen Sinne als Publikations- und Kommunikationsplattform verwenden, bis zu solchen, die dessen mediale Besonderheiten zur eigenen ästhetischen Gestaltung nutzen. Diese Literatur schließt dezidiert an ältere literarische und künstlerische Traditionen der Avantgarde an und tritt mit hohem ästhetischem Anspruch auf, der von Konsekrationsinstanzen des Literaturbetriebs (z. B. dem

¹ Etwas ausführlicher dazu Simone Winko: Literatur und Literaturwissenschaft im digitalen Zeitalter. Ein Überblick. In: Der Deutschunterricht LXVIII/5 (2016), S. 2–13; aus diesem Beitrag übernehme ich im Folgenden einige Passagen.

Deutschen Literaturarchiv Marbach) auch unterstützt wird. Als bekanntestes Beispiel sei hier der schon in die Jahre gekommene *Assoziations-Blaster* von Dragan Espenschied und Alvar Freude genannt (1999), in dem von Nutzern eingegebene Texte von beliebiger Länge miteinander verbunden werden und so ein ständig wachsendes, intern komplex vernetztes Hypertextgebilde entsteht. Neuere technische Möglichkeiten nutzen ‚Texte‘, die Geodaten aus der lebensweltlichen Situation der Nutzer aufnehmen und mit literarischen Mitteln verbinden, etwa mit dem Erzählen einer Geschichte oder lyrischen Formen. So integriert z. B. Stefan Schemats ‚augmented reality fiction‘ *Wasser* (2004) die über GPS gewonnenen Geodaten in eine Erzählung, in der Fiktion und medial vermittelte Wirklichkeit zusammenspielen: Die Geschichte variiert mit den Wegen, die die Rezipienten einschlagen, und welche Wege sie wählen, hängt wiederum mit den Informationen aus der Geschichte zusammen. In Beat Suters und René Bauers literarischer App *AndOrDada* (2012) nimmt die Software Impulse des W-Lan aus der unmittelbaren Umgebung der Leser auf und wandelt sie in ‚poetische Objekte‘ um, wahlweise im narrativen oder lyrischen Modus.

Andere Typen der Netzliteratur sind primär an den sozialen Möglichkeiten des Netzes interessiert. Sie sind Ausdruck der aktuellen Partizipationskultur: User-generierte Inhalte und literarische Beiträge nicht-professioneller Autorinnen und Autoren nehmen im Internet breiten Raum ein. Literatur kann hier zahlreiche Funktionen erfüllen. Als Ausgangsmedium dient sie etwa für literaturbasierte Webserien, die Figuren und Handlungselemente, zum Teil auch Textpassagen, aus kanonischen und damit vielen Lesern bekannten Romanen als Material für einen seriell erscheinenden Video-Blog (Vlog) verwenden. Nutzer dieser Vlogs kommunizieren über die Serie z. B. auf Twitter oder Facebook und produzieren so die öffentliche Anschlusskommunikation, die zu den Merkmalen der Partizipationskultur zählt. Diese literaturbasierten Webserien wiederum sind verwandt mit der noch weiter verbreiteten literarischen Praxis der ‚Fanfiction‘. Auch deren Verfasser nutzen die Figuren und fiktiven Welten vorliegender Werke, indem sie sie modifizieren bzw. als Material verwenden, um eigene Geschichten zu erfinden. Sie bleiben aber meist im textuellen Medium. Ebenfalls ein literarisches Massenphänomen bilden die Schreib- bzw. Literaturforen, die Werkstattcharakter haben und oft literarischen Genres gewidmet sind. Anders als die lebhaft genutzten literarischen Foren, die in der Literaturwissenschaft bislang kaum wahrgenommen wurden,² sind Mitschreibprojekte im Netz etwas stärker in den Fokus der Forschung geraten, ihres Merkmals der kollektiven Autorschaft wegen. Allerdings gibt es nur noch wenige aktuelle Beispiele für dieses Format, ganz im Gegensatz zu den literarischen Blogs. Sie sind in erster Linie durch ihre netzbasierte Software bestimmt, die in ihren Möglichkeiten aber nicht immer ausgeschöpft wird, und können inhaltlich und formal so unterschiedlich

²Eine der Ausnahmen bildet Julian Ingelmann: *Die Wortkrieger des Web 2.0. Laienliterarische Praktiken im Mikrofeld digitaler Schreibforen*. Göttingen 2019.

gestaltet werden, dass es schwierig ist, einen gemeinsamen Nenner anzugeben.³ Schließlich seien noch die spielerischen Varianten genannt: Die Möglichkeiten vernetzter sozialer Interaktion können mit ludischen Elementen und literarischen Genre-Mustern verbunden werden, wie es im deutschsprachigen Raum z. B. das Browserspiel *TwinKomplex* (2011–2014) gezeigt hat.

Literaturwissenschaftliche Forschungen zu Phänomenen wie den gerade skizzierten gibt es seit 30 Jahren. Viele haben sich aber bislang weniger mit den neuen Formen beschäftigt als vielmehr mit den Aspekten der digitalen bzw. Netzliteratur, die sich an alte Debatten im Fach anschließen lassen, und sie haben traditionsreiche und damit etablierte Konzepte herangezogen: so etwa die Inszenierung der Autorpersona im Netz oder die Frage nach der neuen Relevanz der Materialität als Reaktion auf deren ‚digitale Verflüchtigung‘. Eine Beschränkung auf das konzeptuell Bekannte muss nicht immer mit Hinweis auf Beharrungskräfte oder dem Vorwurf der Innovationsverweigerung erklärt werden, sondern hat auch viel mit disziplinären Kompetenzen zu tun. Zwei Erweiterungen dieser Kompetenzen liegen nahe: So wäre, um eine neuere Position aufzunehmen, die erweiterte Kompetenz einer ‚Procedural Literacy‘ zu fordern, die von Literaturwissenschaftlern auch Wissen über ‚allgemeine algorithmische Prinzipien‘ verlangt.⁴ Und Kenntnisse empirischer Forschung sind erforderlich, um nicht bei der allgemeinen Reflexion über Chancen und Risiken des Lesens digitaler Literatur stehenzubleiben, sondern zu erforschen, was sich z. B. im Modus des Lesens ändert, wenn Literatur auf Tablets, Handys oder als enhanced E-Book rezipiert wird.⁵ Um bestimmte Aspekte der neuen literarischen Phänomene zu analysieren und zu erklären, könnten Theorien und Verfahren an ihre Grenzen stoßen, die mit Bezug auf ‚vor-digitale‘ Literatur bzw. Literatur im Sinne der Buchpraktik entwickelt worden sind. Damit ist weder gesagt, dass die neuen Entwicklungen im Gegenstandsbereich Anlass zu einer Generalrevision der literaturwissenschaftlichen Grundbegriffe geben müssen – immerhin gibt es Literatur im traditionellen Sinne in erheblichem Ausmaß auch weiterhin, und sie soll auch noch erforscht werden –, noch soll behauptet werden, eine neue, alles umfassende Großtheorie sei nötig, um die heterogenen Phänomene und ihre Beziehungen zueinander zu modellieren. Im Gegenteil: Nicht der homogenisierende Versuch, mit einer grundlegenden Theorie die Vielfalt aller mit Digitalität verbundenen Aspekte der Literatur – vom Motiv bis hin zur digitalen

³Vgl. dazu Elisabeth Michelbach: Poetik des autobiografischen Blogs. Göttingen 2020.

⁴Martin Stobbe: Quellcode lesen? Ein Plädoyer für Procedural Literacy in den Literaturwissenschaften. In: Sebastian Böck/Julian Ingelmann/Kai Matuszkiewicz/Friederike Schruhl (Hg.): Lesen X.0. Rezeptionsprozesse in der digitalen Gegenwart. Göttingen 2017, S. 47–67, S. 65.

⁵Das Spektrum neuerer Forschungen zu diesem Thema wird entfaltet in Sebastian Böck, Julian Ingelmann, Kai Matuszkiewicz, Friederike Schruhl (Hg.): Lesen X.0. Rezeptionsprozesse in der digitalen Gegenwart. Göttingen 2017. Das Lesen von enhanced E-Books hat Stefanie Lange empirisch untersucht; Stefanie Lange: Enhanced E-Books. Eine empirische Studie zum immersiven Erleben. Stuttgart 2019.

Kommunikationsform – beschreiben und erklären zu können, scheint der Situation angemessen zu sein, gefragt sind vielmehr Revisionen mittlerer Reichweite: In genauen Auseinandersetzungen mit den unterschiedlichen Phänomenen sollte ausgelotet werden, welche Konsequenzen sich aus den Veränderungen des Gegenstandes für die literaturwissenschaftlichen Begriffe und Verfahren ergeben, und welche Implikationen dies für die herangezogenen Theorien hat.

Diesen Herausforderungen stellen sich die Beiträge des Sektionsteils I.1. Das populärkulturelle Phänomen des Blog in einer dezidiert nicht-avantgardistischen Variante behandelt *Jörg Schuster*. Er plädiert für eine Erweiterung des literaturwissenschaftlichen Gegenstands- und Methodenspektrums und untersucht exemplarisch Modeblogs als Beispiele eines spezifischen Umgangs mit Formen- und Sprachmaterial. Dabei führt er zugleich Möglichkeiten eines diskursanalytischen Ansatzes unter Bedingungen digital vorliegender Texte vor. *Julia Nantke* befasst sich mit den Konsequenzen, welche die digitalen Möglichkeiten der Produktion und Rezeption von Literatur für die Auffassungen von ‚Text‘ und ‚Textualität‘ haben, mithin für ein besonders wichtiges literaturwissenschaftliches Begriffsfeld. In ihrer Analyse zweier Beispiele digitaler Literatur – *Tausend Tode schreiben* und *0×0a* – identifiziert sie typische Wechselwirkungen zwischen Prozessen der Dynamisierung und Destabilisierung und solchen der Formgebung und Stabilisierung. Das kollaborative Sammeln und Kommentieren in digitalen Räumen beobachtet *Alexander Nebrig*. Am Beispiel der Plattform „Genius.com“ untersucht er die Annotationspraxis von Laienphilologen und fragt nach Schnittstellen mit der literaturwissenschaftlichen Kommentar- und Interpretationspraxis.

(2) *Literaturwissenschaft unter digitalen Bedingungen*. Unter der zweiten Leitperspektive dieser Sektion zeigte sich, dass es nicht der Gegenstand ‚Literatur‘ in dem gerade skizzierten, dynamischen Sinne zu sein scheint, von dem sich Literaturwissenschaftlerinnen und Literaturwissenschaftler besonders herausgefordert sehen, sondern die unter ästhetisch-medialen Aspekt weniger aufregende Variante, die retro-digitalisierte Literatur. Genauer gesagt interessieren sie besonders die Möglichkeiten, die sich dadurch bieten, dass ein Großteil der Texte vergangener Jahrhunderte in digitaler Form vorliegt. Wie sehen diese Möglichkeiten aus? In welchem Ausmaß ist eine konsequent digitale Literaturwissenschaft wandlungsfähig? Geht es darum, traditionelle Ziele und Verfahren des Faches durch neue Ziele und neue Verfahren zu erweitern oder zu ersetzen? Wenn Gegenstände, Methoden und Praktiken unter digitalen Bedingungen gegebenenfalls anders konzeptualisiert, legitimiert und eingeübt werden – was folgt daraus für die Literaturwissenschaft als Disziplin?

Antworten können nur gefunden werden, wenn man eine globalere Perspektive auf das Fach einnimmt und mögliche oder gewünschte Entwicklungstendenzen zu bestimmen versucht. Dies sollte in dieser Sektion exemplarisch und möglichst nicht im luftleeren Raum von Spekulationen und nicht nur programmatisch geschehen, sondern fundiert durch Beobachtungen im Fach und formuliert mit Bezug auf Herausforderungen, die sich aus der Digitalität von Gegenstand, Verfahren und Praktiken (z. B. Kommunikationswegen) ergeben. Diese Herausforderungen können aus sehr Unterschiedlichem entstehen, z. B. aus der

Konfrontation einer etablierten historisch-hermeneutischen Fachtradition mit digitalen Verfahren, aus Defiziten herkömmlicher textanalytischer, -interpretativer oder historiographischer Verfahren oder aus internen Problemen literaturwissenschaftlicher Theorien, die Ausschlussmechanismen erzeugen.

Michael Stolz lotet die technischen, sozialen und epistemologischen Konsequenzen der Digitalität aus. In Auseinandersetzung mit einem literarischen Schreibexperiment Jean-François Lyotards aus den 1980er Jahren, angereichert durch gegenwärtige Erfahrungen mit digitalen Formaten, identifiziert er fünf miteinander verbundene Themenfelder, in denen sich die Herausforderungen der Digitalität für Literatur und Literaturwissenschaft insbesondere zeigen: die Vernetzung, die digitale Simulation von Wirklichkeit, Kombinatorik und Aleatorik als Verfahren der Textproduktion, künstliche Intelligenz und die Verfügbarkeit von „Wissensvorräten“. Ein Grundproblem der Anwendung empirischer, korpusbasierter Analyseverfahren in der Literaturwissenschaft behandelt *Jonas Kuhn* aus computerlinguistischer Sicht: Lassen sich die Implikationen dieser Verfahren mit essentiellen Annahmen einer an Textinterpretation interessierten historisch-hermeneutischen Literaturwissenschaft überhaupt vereinbaren? Kann die Annahme der statistischen Repräsentativität der Daten – eine Annahme, die bei der linguistischen Anwendung von Computermodellen in der Textanalyse so wichtig wie unproblematisch ist – mit der literaturwissenschaftlichen „Grundmotivation“ verbunden werden, das ästhetisch Besondere eines literarischen Texts oder einer Textgruppe herauszuarbeiten? Im Beitrag wird neben einer „defensiven“ Antwort, die vorschlägt, die Verfahren nur auf bestimmte, unproblematische Texteigenschaften anzuwenden, auch eine weiterreichende „offensive“, die Modellierung als solche betreffende Antwort gegeben.

Der letzte Beitrag dieser Sektion verfolgt methodische und fachgeschichtliche Ziele: *Fotis Jannidis*, *Steffen Martus*, *Leonard Konle* und *Jörn Kreutel* fragen zum einen, wie digitale Daten und Analysemethoden die Erforschung der Fachgeschichte verändern können, und zum anderen, wie sich Kontinuität und Wandel im Fach mithilfe dieser Methoden nachweisen lassen. Sie zeigen an einem Beispiel – dem Korpus der *Deutschen Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* von 1960–2009 –, welche Verfahren für welche fachgeschichtlichen Perspektiven sinnvollerweise eingesetzt werden können und wo methodische Probleme liegen, etwa solche, die sich aus der unterschiedlichen Schärfe von Begriffen in den an Operationalisierbarkeit interessierten Digital Humanities und der an Komplexität ausgerichteten Literaturwissenschaft ergeben. Zugleich zeigt der Beitrag Wege für Anschlussforschungen einer praxeologisch ausgerichteten Wissenschaftsforschung, für die sich der Dualismus ‚empirisch versus hermeneutisch‘ als zu schlicht erweist.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Die Lesbarkeit der Welt. Digitalität als Herausforderung der Literaturwissenschaft

Michael Stolz

1 Vorbemerkungen

Digitalität, verstanden als die „Verwendung digitaler Verfahren und Technologien in der geisteswissenschaftlichen Forschungspraxis“,¹ konfrontiert die Betroffenen mit Herausforderungen, deren Ausmaß zum gegenwärtigen Zeitpunkt noch schwer absehbar ist. Chancen und Risiken liegen in diesem Bereich sehr eng beisammen. Bestandsaufnahmen haben deshalb unter Bedingungen, in denen sich immer mehr Forschende auf die technischen und methodischen Möglichkeiten der Digitalität einlassen und diese damit dynamisch weiterentwickeln, notgedrungen einen vorläufigen Charakter. Dies gilt umso mehr, als sich die Funktionen der Digitalität in einem komplexen Gefüge von „Referenzialität, Gemeinschaftlichkeit und Algorithmizität“ bewegen.² Neben den technischen Verfahren einerseits (Algorithmizität) und den damit bearbeiteten Gegenständen andererseits (Referentialität) prägen sich Rollenmuster und Verhaltensnormen aus, mit denen Nutzer diese Instrumentarien handhaben, auf die jeweiligen Gegenstände zugreifen und miteinander interagieren (Gemeinschaftlichkeit). Gerade solche mit der Entwicklung der Technik und dem objektbezogenen Zugriff verbundenen Interaktionsformen unterliegen in einer Phase sich dynamisch fortentwickelnder Digitalität einer kontinuierlichen Ausdifferenzierung. Digitalität lässt sich dabei definieren als ein „Set von Relationen, das heute auf Basis der Infrastruktur

¹ Vgl. die online publizierte Symposienreihe der DFG: *Digitalität in den Geisteswissenschaften*, <http://digitalitaet-geisteswissenschaften.de/> (letzter Aufruf 14.06.2019).

² Vgl. Felix Stalder, *Kultur der Digitalität*, Berlin 2016, 2; ausführlicher in Kap. 2: „Formen der Digitalität“, 95–202.

M. Stolz (✉)
Universität Bern, Bern, Schweiz
E-Mail: michael.stolz@unibe.ch

digitaler Netzwerke in Produktion, Nutzung und Transformation materieller und immaterieller Güter sowie in der Konstitution und Koordination persönlichen und kollektiven Handelns realisiert wird“.³ Traditionelle Profile wie jene von Produzenten, Nutzern, Rezipienten beginnen sich zunehmend zu überlagern, indem Produzenten von Algorithmen etwa auch deren Nutzer sind, Nutzer von digitalen Daten auch die Rolle von (interpretierenden) Rezipienten übernehmen, Rezipienten der Digitalität aber ihrerseits zu neuen Produzenten werden können. Derartige Interferenzen haben ihren Ursprung in gesellschaftlich-technischen Entwicklungen und wirken wiederum selbst auf gesellschaftliche Prozesse zurück.

Die Literaturwissenschaft partizipiert an solchen technischen und sozialen Entwicklungen und sieht sich mit der Herausforderung konfrontiert, die in ihrem Zuständigkeitsbereich einwickelten Begriffe und Methoden kritisch zu hinterfragen. So haben etwa angesichts der erwähnten Überlagerung von Produzenten-, Nutzer- und Rezipientenrollen konventionelle Sender- und Empfängermodelle nur noch beschränkte Gültigkeit. Aber auch grundlegende Verfahren wie das Sammeln von wissenschaftlichen Daten (*colligere*), deren Auslese und Auswertung (*eligere*) und letztlich die Basistechnik des Lesens (*legere*) unterliegen grundlegenden Modifikationen und erfahren eine neue Einschätzung. Angesichts der Möglichkeit, mithilfe elektronischer Verfahren immer größere Datenmengen zu verarbeiten, steht der Vorgang des Lesens selbst zur Disposition.⁴ Er kann von einem kognitiven individuellen Vorgang an Maschinen ausgelagert werden. Aus dem *Close Reading* als einer extremen Praxis formaler Literaturanalyse ist unter den gegebenen technischen Bedingungen ein *Distant Reading* geworden, dessen diagrammatisch visualisierte Ergebnisse neue Sichtweisen (wenn auch nicht immer grundstürzend neue Erkenntnisse) in Bezug auf literaturwissenschaftliche Gegenstände ermöglichen.⁵ Nicht als ein ‚Lesen auf Distanz‘ sondern als eine Erschließung umfangreicher Textkorpora reichen solche lexikographischen und texthermeneutischen Verfahren bis in die Mitte des vorigen Jahrhunderts zurück, als der Jesuitenpater Roberto Busa damit begann, die Schriften des Thomas von Aquin mithilfe computergestützter Konkordanzen aufzubereiten. Zurecht gilt Busa mit dem von ihm initiierten *Index Thomisticus* inzwischen gemeinhin als der Begründer der *Digital Humanities* als einer wissenschaftlichen Disziplin.⁶

³Ebd., 18.

⁴Vgl. zur Begrifflichkeit Rüdiger Campe, „Lecture“, in: Harald Fricke/Georg Braungart/Klaus Grubmüller u. a. (Hg.), *Realexikon der deutschen Literaturwissenschaft. Neubearbeitung des Realexikons der deutschen Literaturgeschichte*, Bd. 2: H–O, Berlin/New York 2000, 385–388; Hugo Aust, „Lesen“, in: ebd., 406–410; Ursula Rautenberg/Ute Schneider, *Lesen. Ein interdisziplinäres Handbuch*, Berlin/Boston 2015; sowie im vorliegenden Band den Beitrag von Manuel Braun/Sonja Glauch/Florian Kragl, *Unterwegs zum Text ohne Herausgeber und ohne Leser. Eine medienpragmatische Standortbestimmung der digitalen Edition* (Abschn. 2, 295).

⁵Vgl. die Aufsatzsammlung von Franco Moretti, *Distant Reading*, London 2013; ins Dt. übers. von Christine Pries, *Distant Reading*, Konstanz 2016.

⁶Vgl. Thomas N. Winter, „Roberto Busa, S. J., and the Invention of the Machine-Generated Concordance“, in: *The Classical Bulletin* 75/1 (1999), 3–21; Fotis Jannidis/Hubertus Kohle/Malte Rehbein (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart 2017, 3 f.

Der vorliegende Beitrag unternimmt es, die eingangs erwähnte Digitalität angesichts ihrer technischen, sozialen und epistemologischen Konsequenzen als eine ‚Herausforderung der Literaturwissenschaft‘ zu begreifen. Ganz bewusst soll dabei die ‚Lesbarkeit‘ der *litteratura*, das heißt eines (vorwiegend) aus *litterae* – also aus Schriftzeichen – bestehenden textuellen Gewebes im Fokus der Aufmerksamkeit stehen.⁷ Mit diesem Stichwort sind texthermeneutische Verfahren aufgerufen, die seit den Sechzigerjahren im engeren oder weiteren Umfeld der Forschungsgruppe *Poetik und Hermeneutik* entwickelt worden sind und in jüngster Zeit ein neuerliches, durchaus auch kritisch geprägtes Interesse erfahren haben.⁸ Die erwähnten Zugänge konkretisierten sich bekanntlich bei Hans Robert Jaub in einer als *Provokation der Literaturwissenschaft* präsentierten Rezeptionsästhetik (1967),⁹ bei Wolfgang Iser in der Bestimmung des in der Textstruktur angelegten *implizite[n] Leser[s]* und damit einhergehend des *Akts des Lesens* (1972 und 1976),¹⁰ und sie führten bei Hans Blumenberg zur grundsätzlichen Frage nach der *Lesbarkeit der Welt* (1981).¹¹ Diese Konturen sollen im folgenden zweiten Teil des Beitrags mit der gebotenen Kürze in Erinnerung gerufen werden. In einem dritten Abschnitt wird das Konzept der ‚Lesbarkeit‘ sodann mit einem Formexperiment konfrontiert, das einer der führenden französischen Repräsentanten der sogenannten ‚Postmoderne‘, mit der einige jüngere Vertreter der Gruppe *Poetik und Hermeneutik* in produktivem intellektuellen Austausch standen,¹² anlässlich einer Ausstellung am Pariser *Centre Georges Pompidou* durchführgeführt hat: Es handelt sich um die von Jean-François Lyotard (1985) präsentierte Schau *Les Immatériaux*, die sich den kulturellen Herausforderungen

⁷Vgl. Klaus Weimar, „Literatur“, in: Harald Fricke/Braungart/Grubmüller u. a. (Anm. 4), 443–448, hier: 444b.

⁸Vgl. bes. Petra Boden/Rüdiger Zill, (Hg.), *Poetik und Hermeneutik im Rückblick. Interviews mit Beteiligten*, Paderborn 2017; Julia Amslinger, *Eine neue Form von Akademie. ‚Poetik und Hermeneutik‘ – die Anfänge*, Paderborn 2017.

⁹Vgl. bes. die Konstanzer Antrittsvorlesung von Hans Robert Jaub, *Literaturgeschichte als Provokation der Literaturwissenschaft*, Konstanz 1967 (*Konstanzer Universitätsreden* 3); wieder abgedruckt in: Ders., *Literaturgeschichte als Provokation*, Frankfurt a. M. 1970, 144–207 (im Folgenden zitiert). Zur kritischen, v. a. an die NS-Vergangenheit von Jaub geknüpfte Auseinandersetzung vgl. stellvertretend Ottmar Ette, *Der Fall Jauss. Wege des Verstehens in eine Zukunft der Philologie*, Berlin 2016; Jens Westemeier, *Hans Robert Jauss. Jugend, Krieg und Internierung*, Konstanz 2016. Zu Konzepten der Rezeptionsästhetik sei auf die einschlägigen Beiträge in Rainer Warning (Hg.), *Rezeptionsästhetik*, München 1975, verwiesen.

¹⁰Vgl. Wolfgang Iser, *Der implizite Leser. Kommunikationsformen des Romans von Bunyan bis Beckett*, München 1972; Ders., *Der Akt des Lesens. Theorie ästhetischer Wirkung*, München 1976. Dazu kritisch zuletzt: Iris Bäcker, *Der Akt des Lesens – neu gelesen. Zur Bestimmung des Wirkungspotentials von Literatur*, Paderborn 2014.

¹¹Hans Blumenberg, *Die Lesbarkeit der Welt*, Frankfurt a. M. ²1983 [1981] (im Folgenden zitiert). Dazu zuletzt: Thomas Meyer, „Lesbarkeit“, in: *Blumenberg lesen. Ein Glossar*, Berlin 2014, 171–184.

¹²Zu nennen wären etwa Manfred Frank und Rainer Warning; vgl. die Ausführungen in Boden/Zill, *Poetik und Hermeneutik* (Anm. 8), 196, 278–280.

der damals erst in ihren Anfängen befindlichen Computertechnologie und ihrer ‚immateriellen‘ Präsentationsformen widmete.¹³ Als Begleitprojekt inszenierte Lyotard unter dem Titel *Épreuves d'écriture* ein gemeinschaftliches Vorhaben, bei dem 26 französische Intellektuelle zu einschlägigen, mit den epistemischen Herausforderungen des Computerzeitalters in Zusammenhang stehenden Begriffen Stellung nehmen sollten – dies interaktiv in einer medialen, durch den Computer ermöglichten Kooperation.¹⁴ In Auseinandersetzung mit den hier in einem mittlerweile über dreißigjährigen Abstand fassbaren Stellungnahmen und im Blick auf die (schriftlich dokumentierte) Performanz ihrer Interaktionen sollen in einem vierten Teil thesenartig Programmpunkte zu Literatur und Literaturwissenschaft im digitalen Zeitalter formuliert und damit zur Diskussion gestellt werden. Ein kurzer Ausblick soll den Beitrag beschließen.

2 Die Lesbarkeit der Welt

Denn es handelt sich ja nicht darum, die Werke des Schrifttums im Zusammenhang ihrer Zeit darzustellen, sondern in der Zeit, da sie entstanden, die Zeit, die sie erkennt – das ist die unsere – zur Darstellung zu bringen. Damit wird die Literatur ein Organon der Geschichte, und sie dazu – nicht das Schrifttum zum Stoffgebiet der Historie – zu machen, ist die Aufgabe der Literaturgeschichte.

Aus diesem in einer Fußnote zitierten Fazit von Walter Benjamins Aufsatz über *Literaturwissenschaft und Literaturgeschichte* (1931) entwickelt Hans Robert Jauß unter anderem sein Konzept des „Erwartungshorizont[s]“, vor dem sich jede literarische Erfahrung, auch diejenige in einer z. B. zeitlich oder örtlich bedingten hermeneutischer Distanz, vollziehe.¹⁵ Ziel einer „Literaturgeschichte des Lesers“¹⁶ müsse es sein, diesen „Erwartungshorizont zu objektivieren“.¹⁷

¹³Vgl. den Katalog: Jean-François Lyotard (Hg.), *Les Immatériaux. Album et inventaire*, Paris 1985; dazu diverse Texte wie ein Konzept, Interviews und eine Pressemitteilung in: Jean-François Lyotard/Jacques Derrida/François Burkhardt, *Immaterialität und Postmoderne*, aus dem Franz. übers. von Marianne Karbe, Berlin 1985. Zur Aufarbeitung in jüngerer Zeit vgl. Antonia Wunderlich, *Der Philosoph im Museum. Die Ausstellung ‚Les Immatériaux‘ von Jean François Lyotard*, Bielefeld 2008; Yuk Hui/Andreas Broeckmann (Hg.), *30 Years after ‚Les Immatériaux‘. Art, Science, and Theory* [Lüneburg 1915].

¹⁴Vgl. Jean-François Lyotard/Thierry Chaput (Hg.), *Épreuves d'écriture*, Paris 1985. Mitunter wird dieser Band, etwa bei Wunderlich (Anm. 13), 251, als Teil 1 des Ausstellungskatalogs geführt, während *Album et inventaire* als Teile 2 und 3 angegeben werden (aus der Titelei der Katalog erschließt sich dies nicht). Zu weiterführender Literatur vgl. ebenfalls Anm. 13.

¹⁵Nach Walter Benjamin, „Literaturwissenschaft und Literaturgeschichte“, in: Ders., *Ausgewählte Schriften*, Bd. 2: *Angelus Novus*, Frankfurt a. M. 1966 [1931], 450–456, hier: 456; zitiert in Jauß (Anm. 9), 170, Anm. 63, der Begriff des „Erwartungshorizont[s]“ fällt 173.

¹⁶Begriff nach Harald Weinrich, „Für eine Literaturgeschichte des Lesers“, in: *Merkur* 21/236 (1967), 1026–1038; zitiert bei Jauß (Anm. 9), 180.

¹⁷Ebd., 177.

Ermöglicht werde diese Objektivierung durch synchrone Schnitte im Gang durch die Literaturgeschichte, als Stationen „im Vorher und Nachher der Diachronie“.¹⁸ Auf diese Weise lange die literarische Erfahrung auch beim „historischen Standort des gegenwärtigen Betrachters, alias Literaturhistorikers“ an; ebenso müsse der jeweilige „Interpret seine eigene Erfahrung ins Spiel bringen“.¹⁹ Aus dieser die literaturwissenschaftliche Methodik und Literaturrezeption vermengenden Auffassung leitet Jauß schließlich die „gesellschaftliche Funktion der Literatur“ ab, bei der „die literarische Erfahrung des Lesers in den Erwartungshorizont seiner Lebenspraxis eintritt [...] und damit auch auf sein gesellschaftliches Verhalten zurückwirkt“.²⁰ Durch diese am jeweiligen Erwartungshorizont bemessene, fortwährende gesellschaftliche Funktion der Literatur würden auch „Wege zukünftiger Erfahrung eröffnet“.²¹

Die notwendige Voraussetzung eines solchen Prozesses sei, wie Jauß in Anlehnung an Harald Weinrich, wiederum in einer Fußnote, festhält, eine prinzipiell vor dem Hintergrund literatursoziologischer Prämissen in Rechnung zu stellende „Leserrolle, die im Werk implizit enthalten ist“.²² Wolfgang Iser hat dieses Konzept der ‚Leserrolle‘ bekanntlich zum Begriff des ‚impliziten Lesers‘ ausgestaltet, der im literarischen Text entworfen ist. In *Der Akt des Lesens* beschreibt er (in Anlehnung an Roman Ingardens Begriff der ‚Unbestimmtheitsstellen‘) die Funktionen von Leerstelle und Negation: Die Leerstelle bewege sich auf einer syntagmatischen Achse der Textstrategie, indem der Leser bei seiner Interaktion mit dem Text dazu aufgerufen werde, einzelne Textsegmente und Darstellungsperspektiven miteinander zu kombinieren. Mittels seiner Vorstellungskraft fülle der Leser die vom Text offen gelassenen Lücken. Er favorisiere dabei einzelne in den Textsegmenten angelegte Positionen, mache sie zum „Thema“, während er andere marginalisiert, indem er ihnen den „Charakter des Horizonts“ zuweise.²³ Auf diese Weise werde die „Struktur der Leerstelle“ zum „Konstitutionsmodus des Textes im Lesevorgang“ und „ermöglicht [...] die Beteiligung des Lesers am Vollzug des Textgeschehens“.²⁴ Leerstellen auf einer paradigmatischen Achse der Textstrategie fasst Iser unter dem Begriff der ‚Negation‘. Hierbei handele es sich um ein in Texten bspw. im Rahmen von Tugendkatalogen aufgerufenes Normenrepertoire, das den Handlungserfolg einer Figur – angesichts der Konfrontation mit einer korrupten Umwelt – nicht gewährleisten könne und deshalb im Vollzug der Lektüre ‚negiert‘ werde (Iser nennt den Protagonisten Abraham Adams in Henry Fieldings Roman *Joseph Andrews*

¹⁸ Ebd., 197.

¹⁹ Zitate ebd., 191 f.

²⁰ Ebd., 199.

²¹ Ebd., 202.

²² Ebd., 181, Anm. 85, nach Weinrich (Anm. 16).

²³ Vgl. die Begriffe bei Iser (Anm. 10 [1976]), 306 f.

²⁴ Ebd., 314.

als Beispiel).²⁵ Während die auf einer syntagmatischen Achse organisierte Leerstelle einer Textstrategie der „Ausparung“ unterliege, werde die auf einer paradigmatischen Achse organisierte Negation mittels einer Textstrategie der „Aufhebung“ erreicht.²⁶ In beiden Fällen werde dabei „der formulierte Text durch Unformuliertes gedoppelt“ – eine Doppelung, die Iser „als die Negativität fiktionaler Texte bezeichnet“.²⁷ Die so gefasste Negativität zeichne sich dadurch aus, dass sie „als das Unformulierte ein noch Unbegriffenes markiert“ und „dadurch den Leser in eine Beziehung zum Text“ setze; Negativität ermögliche es auf diese Weise, „unser Mittendrinsein im Leben“ bzw. „die Welt zu transzendieren“.²⁸ Die von Jauß veranschlagte „gesellschaftliche Funktion der Literatur“ (s. o.) weicht damit einem Akt des Lesens, der, indem er die Lebenspraxis übersteige, sich dieser gerade enthebe. Der so verstandene Akt des Lesens lässt Raum für eine „Aleatorik“, in der „die Kombinationsmöglichkeiten nicht festgelegt“ sind; er erweist, dass es „den einen Sinn fiktionaler Texte nicht gibt“, dass vielmehr diese Ermangelung „in den verschiedensten Kontexten immer wieder Sinn zu geben vermag“.²⁹

Hans Blumenberg bezieht diesen ‚Akt des Lesens‘ auf das – über die verschiedenen historischen Epochen seit der Bibel und den Kirchenvätern bis in die eigene Gegenwart nachweisbare – Konzept von der „Lesbarkeit der Welt“: der Auffassung, die Welt sei als Text oder Buch zu lesen. Literarische Erfahrung, wie sie Jauß und Iser als hermeneutischen Prozess und Transzendenz des Lebens formulieren, wird dabei mit der Erfahrung von Welt konfrontiert, dies unter dem ausdrücklichen Vorbehalt, dass eine solche Erfahrung auch unter dem Verdacht einer „Abmagerung“ gelebter Erfahrung stehen könne.³⁰ Blumenberg kommt dabei von seinen metaphorologischen Ansätzen her; er fasst ‚Lesbarkeit‘ als eine „Metapher für das Ganze der Erfahrbarkeit“.³¹ „Bücherwelt und Weltbuch“ stehen in einem permanenten Spannungsverhältnis: „Zwischen den Büchern und der Wirklichkeit ist eine alte Feindschaft gesetzt. [...] Die geschriebene und schließlich gedruckte Tradition ist immer wieder zur Schwächung von Authentizität der Erfahrung geworden“.³² Diese in der *Lesbarkeit der Welt* früh formulierte Auffassung geht durchaus auch mit einer kritischen Sicht auf die von Jauß und Iser im Blick auf die Rezeption und den Akt des Lesens pointierte literarische Erfahrung einher: „Erfahrung wird angelesen, aber gerade nicht die eigene“.³³

²⁵ Ebd., 327–331.

²⁶ Ebd., 348.

²⁷ Ebd.

²⁸ Ebd., 354.

²⁹ Ebd., 355.

³⁰ Vgl. den Paratext „Über dieses Buch“ in Blumenberg (Anm. 11), unpaginiert [3].

³¹ Überschrift des ersten Kapitels ebd., 9.

³² Überschrift des zweiten Kapitels und dessen Einleitung ebd., 17.

³³ Paratext (wie Anm. 30). Zu den Ambivalenzen einer Hermeneutik, die „ihrem Gegenstand [unterstellt], sich durch ständig neue Auslegung anzureichern“, ebd., 21. In Blumenbergs Studie über *Die Arbeit am Mythos*, Frankfurt a. M. 62001 [1979], 94–96, findet sich eine explizite Kritik

Blumenberg verfolgt die von der Buchmetapher geleistete Herstellung der Totalität von Erfahrung an verschiedenen geistesgeschichtlichen Stationen, von denen einige hier im Blick auf die in den folgenden Kapiteln darzulegenden Phänomene von ‚Digitalität‘ herausgegriffen werden sollen. Dabei lässt sich vorab daran erinnern, dass die in der Informatik übliche Kodierung mit zahlensymbolischen Chiffren in der Begrifflichkeit von *Digits* ihrerseits in metaphorischer Verwandtschaft mit einer altchristlichen Buch- bzw. Schriftmetapher steht: ‚Die Welt ist wie ein von Gottes Finger geschriebenes Buch‘ („universus enim mundus [...] quasi quidam liber scriptus est digito Dei“).³⁴ Der Satz wird unter dem Namen des in Paris wirkenden Mönchs Hugo von St. Viktor tradiert.³⁵ Das darin dokumentierte Analogieverhältnis von der ‚Welt als Buch‘ (Erschaffen als Schreiben) geht auf die griechisch-römische Deutung des biblischen Heilsgeschehens zurück, die das Christentum seit seinen Anfängen prägte und die es im wirkmächtigen Einfluss der Patristik zu einer Buchreligion machte. Das ‚Buch der (heiligen) Schrift‘, das ‚Buch des Lebens‘ bzw. das ‚Buch der Leben‘ sind Metaphern, die in dieser Frühzeit entwickelt und von Kirchenvätern wie Augustinus tradiert wurden.³⁶ Doch auch die (kosmische) Natur wurde mit der Buchmetapher gefasst – als zweites Buch, welches neben das ‚Buch der Schrift‘ trat. Für dieses Konzept steht das Hugo von St. Viktor zugeschriebene Zitat. Es dokumentiert beispielhaft eine seit dem Hochmittelalter in den Vordergrund tretende Auffassung, der zufolge die Natur aus dem Gedanken Gottes hervorgegangen sei und folglich „nicht durch Anschauung, sondern durch Denken verstanden, nämlich ‚gelesen‘ werden kann“.³⁷

Mit der Säkularisierung der Neuzeit, die sich bereits im mittelalterlichen Aristotelismus und den spätmittelalterlichen Laienbewegungen anbahnt,³⁸ löst sich die Natur aus dem Zugriff der göttlichen Schreibhand. Das 17. Jahrhundert sucht eine Weltformel zu ergründen und entwirft mit Leibniz’ *Ars combinatoria* das „Programm einer exakten und vollständigen Darstellung aller gedanklichen Möglichkeiten“.³⁹ Damit einher geht die Idee einer „Universalbibliothek aller

an Isers Konzepten des ‚impliziten Lesers‘ und der „Leerformen mit Besetzungssignalen“, die als „Wiederkehr des schöpferischen Subjekts auf der anderen Seite, auf der der Rezeption“, qualifiziert werden (95).

³⁴ Hugo von Sankt Viktor, „Eruditionis Didascalicae Libri Septem“, in: *Migne, Patrologia Latina* 176 (1854), 739–838, hier: 814 B.

³⁵ Blumenberg (Anm. 11), 51 f., führt ihn als einen Beleg der „eher begeisterten als gelehrten Autoren“ des 12. Jahrhunderts an.

³⁶ Vgl. ebd., 24–29.

³⁷ Ebd., 48.

³⁸ Vgl. ebd., 54–56 u. 58–67.

³⁹ Ebd., 129.

durch Kombinatorik erzeugbaren Bücher“;⁴⁰ eine solche ideale „kombinatorische Bibliothek hätte keinen Leser“.⁴¹ Später mündet diese Auffassung in die „imaginäre Totalintelligenz von Laplace“, der am Übergang vom 18. zum 19. Jahrhundert ein Konzept des Determinismus entwirft, das „von jedem beliebigen Gesamtzustand der Welt eines gegebenen Zeitpunkts her den Gesamtzustand zu jedem beliebigen anderen Zeitpunkt, gleichgültig ob in Vergangenheit oder Zukunft zum gegebenen Zeitpunkt liegend, kalkulatorisch bestimmen“ könne.⁴² Als ‚Laplace’scher Dämon‘ findet dieses Konzept Verbreitung. Der in solcher Kombinatorik und Kalkulatorik manifestierte Universalanspruch konkretisiert sich buchstäblich in den voluminösen Bänden der Enzyklopädien des 18. Jahrhunderts, namentlich in jener von Diderot und d’Alembert: „Wie sähe das dem absoluten Aspekt äquivalente System aus? Es wäre ein ungeheuerliches Buch (*un volume immense*) [...]. Die vollkommene, nämlich aperspektivische Enzyklopädie wäre die Verdoppelung der Natur selbst, das ‚Buch von der Natur‘ in Idealkonkurrenz mit dem ‚Buch der Natur“.⁴³ Der Artikel „Encyclopédie“ in der *Encyclopédie* von Diderot und d’Alembert sieht angesichts der stetig anwachsenden Masse von Büchern einen Moment vorher, „wo es nahezu ebenso schwierig sein werde, sich eine Auskunft aus einer Bibliothek zu beschaffen, wie an der Welt selbst“.⁴⁴

Unter den von Blumenberg angeführten Stationen sei hier – bei verwegener Aussparung großer Epochen wie der Klassik, Romantik und Frühmoderne – nur noch das Schlusskapitel der Abhandlung berührt, in dem der „genetische Code und seine Leser“ zum Thema werden:⁴⁵ Im Blick auf Erwin Schrödingers Konzept des ‚aperiodischen Kristalls‘ und dessen Überführung in die elegante Struktur einer Doppelhelix durch James Watson und Francis Crick betont Blumenberg erneut den metaphorischen Charakter des gewählten Modells: „Die Komplementarität der Nukleotide Adenin und Thymin einerseits, Guanin und Cytosin andererseits ging unter dem Begriff der Paarigkeit der Basen, des *base-pairing*, in die ersten Modellvorstellungen von der Doppelsträngigkeit der DNA ein“.⁴⁶ Die Reproduktion der DNA fassen die Biochemiker in der Terminologie der Buchschriftlichkeit, es

⁴⁰ Ebd., 135.

⁴¹ Ebd., 133.

⁴² Ebd., 138.

⁴³ Ebd., 167.

⁴⁴ Ebd., 167, nach Denis Diderot/d’Alembert [d. i. Jean-Baptiste le Rond], *Encyclopédie, ou Dictionnaire raisonné des Sciences, des Arts et des Métiers*, Paris 1751–1780, Bd. 5, Paris 1755, 644D:

Tandis que les siècles s’écoulent, la masse des ouvrages s’accroît sans cesse, et l’on prévoit un moment où il seroit presque aussi difficile de s’instruire dans une bibliothèque, que dans l’univers, et presque aussi court de chercher une vérité subsistante dans la nature, qu’égagée dans une multitude immense de volumes.

⁴⁵ Ebd., 372–409.

⁴⁶ Ebd., 382.

wird „abgelesen, kopiert oder übersetzt“.⁴⁷ Aber ist es möglich, „den genetischen Code seiner metaphorischen Herkunft zu entziehen“? – fragt Blumenberg.⁴⁸ „Wenn Schrödinger die Gesamtheit der genetischen Faktoren im Zellkern als codierte Schrift ansah, wer war dann der metaphorisch unvermeidliche Leser ihres Textes?“⁴⁹ Schrödingers Rückgriff auf den Laplace’schen Dämon mitsamt dessen universalem Determinismus erscheint Blumenberg an dieser Stelle nicht überzeugend, ja prekär. So führt der beschrittene Weg durch die ‚Lesbarkeit der Welt‘ angesichts des Modells vom ‚genetischen Code‘ ins Offene: Wird letzterer seines metaphorischen Status entkleidet, muss es als Korrektiv zur ‚Verderblichkeit der Texte‘ (sc. des genetischen Codes) einen der negativen Entropie entgegenwirkenden ‚sortierenden Dämon‘ geben: „Seine *triage* verbindet Lesen im Konzept und Auslesen im Substrat“.⁵⁰ Die ‚Lesbarkeit der Welt‘ birgt in der Ermächtigung der naturwissenschaftlichen Erkenntnis das Potenzial der Auslese, der Selektion. Der ‚genetische Code‘ als Metapher bemäntelt in einem illusorischen ‚Als-ob‘ die wissenschaftliche Vereinnahmung der Natur angesichts der Möglichkeit einer künstlichen Züchtung des Erbguts – so suggerieren es Ausführungen auf den letzten Seiten von Blumenbergs Abhandlung: „[E]s war, als sperrte sich die Natur gegen die Endgültigkeit ihrer Verwissenschaftlichung, gegen die ‚Reinheit‘ der Theorie als das Kriterium ihrer Autonomie“.⁵¹

Blumenbergs Entwurf einer *Lesbarkeit der Welt* fasst das ‚Buch der Natur‘ als eine Ermöglichung der Welterfahrung im Zeichen der Schrift, welche den Zugriff auf die Welt zugleich gewährt und verstellt. Es bleibt – wie beim genetischen Code – eine ‚Lücke zwischen Metapher und Modell‘;⁵² es bleibt ganz grundsätzlich eine Kluft zwischen Erlesen einerseits und Erfahren andererseits, zwischen ‚Bücherwelt und Weltbuch‘, zwischen ‚den Büchern und der Wirklichkeit‘.⁵³ Diese Spannung aber potenziert sich, wenn die Bücherwelt zu einer digitalen Welt mutiert. Die damit verbundenen Konsequenzen für das Lesen und dessen Erforschung, die Wissenschaft der *litterae*, soll im Folgenden anhand eines mittlerweile ebenfalls historischen Fallbeispiels betrachtet werden, bei dem die Schrift buchstäblich ‚auf den Prüfstand‘ geraten ist.

⁴⁷ Ebd., 384.

⁴⁸ Ebd., 376.

⁴⁹ Ebd.

⁵⁰ Zitate ebd., 404 u. 407.

⁵¹ Ebd., 408.

⁵² Ebd., 376.

⁵³ Ebd., 16.

3 Épreuves d'écriture

Im Jahr 1985 kuratierte der französische Philosoph Jean-François Lyotard gemeinsam mit Thierry Chaput und Philippe Délis die Pariser Ausstellung *Les Immatériaux* (*Centre Georges Pompidou*, 28. März bis 15. Juli 1985). Sie hatte das Ziel, den von der neuen Telekommunikation und Informatik ausgelösten Wandel in den gesellschaftlichen Strukturen aufzuzeigen. Das Publikum gelangte auf fünf Wegen durch einen Parcours von insgesamt 61 Stationen und wurde dabei mit Bereichen des alltäglichen Lebens und der Ästhetik konfrontiert (z. B. Schachspiel, Bekleidungsstoffe als ‚zweite Haut‘, Lichtästhetik, ‚Selbsterzeugung‘ durch einen Roboter).⁵⁴ Über Installationen wie Bild- und Filmprojektionen, Computerterminals und multimediale Einrichtungen erfolgte eine regelrechte Immersion in die damals neuen Technologien. Eine systematische Lenkung der Besucher war dabei nicht beabsichtigt. Vielmehr trat Lyotard „im Duktus eines philosophischen Lehrers (auf), dessen zentrales Anliegen es ist, seine Schüler so weit zu verunsichern, dass sie einer eigenen Auseinandersetzung nicht mehr ausweichen können“.⁵⁵

Der Titel *Les Immatériaux* bezog sich auf die stoffliche Eigenart, die elektronische Daten gegenüber der Materie traditioneller Kommunikationsformen aufweisen: „eine Struktur, in der der herkömmliche Gegensatz zwischen Geist und Materie keinen Platz mehr hat“.⁵⁶ Konsequenter spielte Lyotard dabei mit der in dem Wort ‚Materie‘ enthaltenen indogermanischen Wurzel „mât“, die ‚mit der Hand machen, messen, bauen‘ bedeutet.⁵⁷ Er teilte jedem der oben erwähnten fünf Wege einen mit dieser Wurzel in Zusammenhang stehenden Begriff zu, wobei er die Semantik dieser Begriffe auf Kommunikationsmodelle, wie jene von Harold D. Lasswell und Roman Jakobson, bezog.⁵⁸ Auf diese Weise ergab sich eine Begriffskonstellation von *maternité* (Sender) – *materiel* (Empfänger) – *matrice* (Code) – *matériau* (Kanal) – *matière* (Referent). Aus der Kombination dieser Begriffe resultierte ein weiterer Begriff, der als ‚immaterielles‘ Dazwischen jedoch keinem der fünf Wege zugeordnet war: *message* (Nachricht).⁵⁹

⁵⁴Vgl. unter der in Anm. 13 genannten Literatur bes. die (jeweils ohne Autorangabe) geführten Beiträge in: Lyotard/Derrida/Burkhardt (ebd.), „Immaterialien. Presse-Mitteilung vom 8. Januar 1985“, 7–18, und „Immaterialien. Konzeption“, 75–89.

⁵⁵Wunderlich (Anm. 13), 9. Vgl. zu den oben erläuterten Charakteristika der Ausstellung ebd., 9–12.

⁵⁶„Philosophie in der Diaspora. Gespräch zwischen Jacques Derrida und Jean-François Lyotard“, in: Lyotard/Derrida/Burkhardt (Anm. 13), 19–26, hier: 23.

⁵⁷Vgl. „Immaterialien. Konzeption“ (Anm. 54), 78.

⁵⁸Vgl. Harold D. Lasswell, „The Structure and Function of Communication in Society“, in: Lyman Bryson (Hg.), *The Communication of Ideas. A Series of Addresses*, New York 1948, 37–51; Roman Jakobson, „Closing Statement: Linguistics and Poetics“, in: Thomas A. Sebeok (Hg.), *Style in Language*, Cambridge, MA, 1960, 350–377.

⁵⁹Vgl. „Immaterialien. Konzeption“ (Anm. 54), 82 f.; Wunderlich (Anm. 13), 51–54.

Das Ausstellungskonzept lässt in seiner Eigenart Theoreme erkennen, mit denen sich Lyotard seit den späten Siebzigerjahren einen Namen als Philosoph der ‚Postmoderne‘ gemacht hatte. In *La Condition postmoderne* (1979) hatte er das Ende der *grands récits* (der großen sinnstiftenden Erzählungen der Moderne wie jener der Aufklärung oder der hegelianisch-marxistischen Geschichtsphilosophie) erklärt.⁶⁰ In dem kurz vor der Ausstellung erschienenen Werk *Le différend* (1983) war der ‚Widerstreit‘ inkommensurabler Sprachspiele Thema: Eine übergeordnete Instanz, die einen solchen ‚Widerstreit‘ schlichten oder auch nur reglementieren könnte, gebe es nicht mehr.⁶¹ Die Ausstellung *Les Immatériaux* konfrontierte diese Ansätze mit dem Einbruch der neuen Technologien in die ‚postmoderne‘ Gesellschaft. Auf „halbfrei zu wählen(den)“ Wegen, „nicht gezwungen, aber doch geleitet“,⁶² sollten sich die Besucher den medialen Herausforderungen der Gegenwart stellen. Die „mât“-Begriffe wiesen dabei in ihrem Wechselspiel auf eine neue, noch unbekanntere Welt.

Bestandteil des Ausstellungskonzepts war ein Schreibexperiment, das Lyotard unter dem Titel *Épreuves d'écriture* durchführte. Es ist unter dem gleichnamigen Titel als Begleitband zur Ausstellung dokumentiert.⁶³ Vor Beginn der Ausstellung beauftragte Lyotard 26 französische Intellektuelle, sich an der Erstellung eines „petit lexique des immatériaux“ zu beteiligen.⁶⁴ In den Monaten September bis Dezember 1984 waren Schriftsteller, Wissenschaftler, Künstler, Philosophen und Linguisten aufgefordert, Begriffe, die im engeren oder weiteren Zusammenhang mit dem Kommunikationsmodell der Ausstellung und den neuen Technologien standen, zu erläutern – dies an dem für die meisten von ihnen noch ungewohnten *Personal Computer*. Die Teilnehmenden bezogen zum individuellen Gebrauch ein Gerät des Typs *Olivetti M20*, auf dem sie ihre Beiträge verfassen und über eine Postverbindung an einen Zentralcomputer (Typ *Olivetti M24*) senden konnten. Dort wurden die Aufzeichnungen gesammelt und allen anderen Beiträgern zur Verfügung gestellt, so dass diese reagieren und ein interaktiver Austausch entstehen konnte. Lyotard ging es dabei um nicht weniger als eine Inszenierung seines Konzepts des ‚Widerstreits‘ unter den neuen technologischen Bedingungen: „Il fallait que l’atelier des divergences devînt un laboratoire des différends“.⁶⁵

Zugleich sollten dabei die Möglichkeiten eines Mediums erprobt werden, das Lyotards Kollege Jacques Derrida, der seinerseits an dem Experiment teilnahm, wenige Jahre zuvor zum zeichentheoretischen Leitbegriff einer damals noch

⁶⁰ Jean-François Lyotard, *La Condition postmoderne. Rapport sur le savoir*, Paris 1979.

⁶¹ Jean-François Lyotard, *Le différend*, Paris 1983.

⁶² Lyotard/Derrida/Burkhardt (Anm. 13), „Immaterialien. Presse-Mitteilung vom 8. Januar 1985“, 12.

⁶³ Vgl. Lyotard/Chaput (Anm. 14). Dazu Wunderlich (Anm. 13), 58–63; Yuk Hui/Andreas Broeckmann (Hg.), „Introduction“, in: Dies. (Anm. 13), 9–24, hier: 16–18.

⁶⁴ Vorwort in: Lyotard/Chaput (Anm. 14), 6. Vgl. zum Folgenden ebd., 6 f., und Lyotard/Derrida/Burkhardt (Anm. 13), „Immaterialien. Presse-Mitteilung vom 8. Januar 1985“, 16–18.

⁶⁵ Lyotard/Chaput (Anm. 14), 6.

analogen Kultur erhoben hatte: die ‚Schrift‘. Der Titel *Les Épreuves d'écriture* lässt sich unter diesen Vorzeichen als durchaus mehrdeutiger Begriff lesen: als ‚Prüfungen‘ oder, wie oben vorgeschlagen, als ‚Prüfstand der Schrift‘. Zugleich handelt es sich um ‚Schriftproben‘, die in dem neuen Medium der Informatik abgegeben werden. Eine besondere Paradoxie besteht dabei darin, dass der französische Ausdruck „épreuve“ auch einen ‚Probe-‘ oder ‚Korrekturabdruck‘ bezeichnet und damit auf das alte, damals noch konventionelle Medium der Drucktechnik verweist. Diese Paradoxie ist in dem Begleitband zudem auch performativ fassbar, denn die ‚Lexikoneinträge‘ und der im Zuge ihrer interaktiven Anfertigung dank der elektronischen Technologie erfolgte flexible Austausch sind ihrerseits in einer erstarrten Schriftform, dem in Druck vorgelegten Begleitband, dokumentiert. Die Reaktion und Rezeptionsformen der Ausstellungsbesucher, welche die Texte der *Épreuves d'écriture* an den Bildschirmen im *Centre Pompidou* lesen konnten, ist erst gar nicht erhalten. Eine solche inadäquate Überlieferungsform erinnert an vergleichbare Paradoxien etwa in der handschriftlichen Tradierung mittelalterlicher Lyrik, die sich, oftmals ohne Melodien und erst recht ohne den Kontext der Aufführung unter körperlich Anwesenden, nur im Medium der Schrift erhalten hat.⁶⁶

Für das angestellte ‚Schriftexperiment‘ hat Lyotard feste Regeln vorgegeben.⁶⁷ Aus der Liste von 50 Wörtern im Umfeld der Problematik von *Les Immatériaux* mussten die Teilnehmenden mindestens 15 bis 20 in Einträgen von zwei bis zehn Zeilen bearbeiten (eine Zeichenzahl, wie in der gegenwärtigen digitalen Kommunikation üblich, war nicht vorgegeben). Da die Teilnehmenden über den Zentralcomputer am *Centre Pompidou* miteinander verbunden waren, sollten sie die Lektüre der fremden Einträge zur Revision ihrer eigenen Beiträge nutzen, aber auch auf die fremden Einträge, deren Revision und Diskussion reagieren. Sie waren eigens dazu aufgerufen, die aus dieser Situation resultierende Erfahrung eines kollektiven Schreibens zu kommentieren.⁶⁸ Das Experiment hatte also auch eine Metaebene, in der die Bedingungen des Schreibens im neuen Medium und die dabei ermöglichten sozialen Prozesse reflektiert werden sollten.

Das Resultat dieses ‚Widerstreits‘ ist im Begleitband *Les Épreuves d'écriture* in der Weise dokumentiert, dass die Beiträge der einzelnen Artikel mit einem Namenkürzel aufgeführt sind; eine ergänzende Ziffer zählt sämtliche Beiträge der Beteiligten durch. Auf diese Weise wird es ermöglicht, die einzelnen Stellungnahmen exakt zu identifizieren und mit der Formel „X répond à Y“ miteinander in Beziehung zu setzen (so in einer Liste am Beginn eines jeden Artikels). An den

⁶⁶Vgl. zur Problematik Paul Zumthor, *La Lettre et la voix. De la ‚littérature‘ médiévale*, Paris 1987; Jan-Dirk Müller (Hg.), *‚Aufführung‘ und ‚Schrift‘ in Mittelalter und Früher Neuzeit*, Stuttgart/Weimar 1996.

⁶⁷Vgl. Lyotard/Chaput (Anm. 14), 6 f.

⁶⁸Vgl. ebd., 7: „nous souhaitons, en particulier, que vous commentiez les modifications que cette situation entraîne dans votre expérience d'écriture“.

Rändern der einzelnen Beiträge stehen Namenkürzel und Ziffern, ergänzt durch eine Datumsangabe (linker Rand), und Verweise auf die Antworten (rechter Rand).

Zu den 26 Autoren gehören neben Derrida etwa der Maler Daniel Buren, der Schriftsteller Michel Butor, der Soziologe Bruno Latour oder der Anthropologe und Linguist Dan Sperber; auffällig ist aus gegenwärtiger Sicht die geringe Zahl an beteiligten Frauen, die mit den Philosophinnen Christine Buci-Glucksmann und Isabelle Stengers vertreten sind. Zu den von diesem Kollektiv bearbeiteten Begriffen zählen einschlägige Termini wie *code*, *dématisation*, *écriture*, *espace*, *langage*, *métamorphose* oder *réseau*, die teilweise auch in Kombination begegnen; auf einige dieser Stichwörter wird im Folgenden zurückzukommen sein.

In einem Nachwort zieht Lyotard zusammen mit einigen Mitarbeiterinnen Bilanz.⁶⁹ Dabei wird deutlich, dass es den Initiatoren vor allem darum gegangen war, die angesichts der neuen Technologie manifeste Änderung des Schreibens zu beobachten („une altération ambiante de l’activité d’écrire“).⁷⁰ Das Nachwort, welches das Privileg eines übergeordneten Beobachterstandpunkts besitzt, versteht sich dabei mit einer gewissen Bescheidenheit als Teil der vorausgehenden, im Rahmen des Experiments entstandenen Texte. Daraus legitimiert sich auch, dass es in die Betrachtungen des folgenden Kapitels entsprechend integriert werden kann. Die vier Hauptabschnitte des Nachworts seien im Folgenden kurz zusammengefasst:

Unter dem Doppelstichwort „matériau, matériel“ werden Leitbegriffe des der Ausstellung zugrundeliegenden Kommunikationsmodells aufgegriffen, hier jedoch bezogen auf das Medium und den Kanal der elektronischen Kommunikation.⁷¹ Angesichts der neuen Technologie seien die Beiträger in die Rolle von ‚Novizen‘, ja von Kindern versetzt worden („infantilisation“). Die mit anderen Computern vernetzte Arbeit am Bildschirm setze den Text sofort in Umlauf („on place déjà le texte dans le réseau de sa circulation“); die mühsame Arbeit des „écrivain-artiste“ auf dem Papier mit den zugehörigen Schreibspuren von Streichungen, Korrekturen, Einfügungen sei nicht mehr fassbar.⁷²

Aber auch der Empfänger („destinataire“) solcher Nachrichten sei eine unbestimmbare, zumindest nicht im traditionellen Sinne bestimmbar Größe.⁷³ In dem neu entstandenen Bezugsnetz des Austauschs von Botschaften werden die Leser ihrerseits zu Co-Autoren, die nicht mehr daran glaubten, dass ein solcher Schreibprozess je an sein Ende gelangen könnte – und damit im traditionellen Sinne ‚lesbar‘ („lisible“) sei. Doch während der Autor sich noch der Illusion hingeben könne, immerzu nur sein eigener Leser zu sein, verschwinde die Privatheit

⁶⁹ Ebd., 259–263. Vgl. zum Folgenden auch Wunderlich (Anm. 13), 60–62.

⁷⁰ Lyotard/Chaput (Anm. 14), 259.

⁷¹ Ebd., 259 f.

⁷² Ebd.

⁷³ Ebd. 261.

des Schreibens im netzwerkartigen Gefüge der Co-Autorschaft: „Le public envahit le privé, la singularité du contenu se dissipe sous le regard collectif“.⁷⁴

Damit stünden der Status von Autorschaft und ‚Autorität‘ („auteur“ – „son statut propre, l’autorité“) zur Disposition.⁷⁵ Der Autor scheine seine Herrschaft über das von ihm Geschriebene zu verlieren („La maîtrise de l’auteur sur son écrit paraît illusoire“); die neue Situation des Schreibens habe zu Unvorhergesehenem („l’aléa y est inclus comme un risque à courir“) und Unordnung („désordres“) geführt.⁷⁶ Erfahrene Beteiligte seien mit dieser Situation souverän, ja spielerisch umgegangen und hätten ihre gewohnte Schreibart beibehalten; doch hätten nicht alle Beteiligten die vorgegebenen Regeln aufrichtig befolgt, einige hätten sich zurückgezogen.

Was die Schrift („écriture“) betreffe, so bleibe ihr Status offen:⁷⁷ „à qui, à quoi est due l’écriture?“ – einige Autoren bedienten sich der Schrift („se servir“), andere stellten sich in ihren Dienst („se trouver à son service“).⁷⁸ Unter den Bedingungen der neuen Technologien bedeute dies, dass die in der Schrift übermittelten Botschaften transparenter würden, die Sprache zugleich einfacher und klarer. Umgekehrt müssten die elektronischen Sprachmaschinen, um es mit den vormaschinellen Sprachen aufnehmen zu können, komplexer werden. Die Funktion der Schrift („le service de l’écriture“) – und damit, so wird man folgern dürfen, die Leistungsfähigkeit einer Schriftkultur – erwiesen sich gerade in ihrem Vermögen zur Komplexität. Der Prüfstand, auf den Lyotard die ins digitale Medium übertragene Schrift mit dem Experiment der *Épreuves d’écriture* gestellt hat, führt damit zu einem insgesamt optimistischen Ergebnis.⁷⁹

Gegenüber den hermeneutischen Ansätzen aus dem Umfeld der Gruppe *Poetik und Hermeneutik*, wie sie eingangs mit der literarischen Erfahrung der ‚Rezeption‘ (Jauß), der Lektüre von ‚Leerstellen‘ (Iser) und der metaphorischen Tragweite einer ‚Lesbarkeit der Welt‘ (Blumenberg) vorgestellt wurden, zeigt sich hier der Neuansatz von Lyotards Erprobung der Schrift: Ihr Gebrauch im Kontext der ‚Immaterialien‘ neuer Technologien bringt Medium, Sender und Empfänger der ausgetauschten Botschaften in ein neues Verhältnis und verändert deren Status. Letztlich wird dabei auch die Rolle der Schrift als Instrument und Determinante neuen Funktionen zugeführt. Mit der historischen Distanz von über dreißig Jahren

⁷⁴ Ebd.

⁷⁵ Ebd., 261 f.

⁷⁶ Ebd., 262.

⁷⁷ Ebd., 262 f.

⁷⁸ Ebd., 263.

⁷⁹ Lyotards allzu (schrift-)emphatisches und letztlich anti-humanistisches Schlusswort muss man dabei nicht zwangsläufig teilen; er folgert, dass sich das mit der Schrift verbundene Vermögen zur Komplexität nicht menschlicher Leistungskraft verdanke, sondern diese vielmehr dem der Schrift eigenen Vermögen zur Komplexität geschuldet sei („L’écriture [...] comme la tâche de tout compliquer, n’est peut-être pas due aux humains, mais ceux-ci dûs à cette tâche“; ebd., 263).

ermöglicht das von Lyotard initiierte Schreibexperiment einen Blick auf die Bedingungen der Schrift und ihrer Lesbarkeit angesichts der Transformation in das digitale Medium. Vergangene literatur- und kulturwissenschaftliche Avantgarden wie die von Jauß angeregte Rezeptionsästhetik oder der von Lyotard inszenierte ‚Widerstreit‘ im elektronischen Netzwerk der ‚Immatériaux‘ bergen das Potenzial als prämaturierte Beschreibungen von Bedingungen der Digitalität gelesen werden zu können. Im Dialog mit den *Épreuves d’écriture* und den ihnen vorausgehenden Ansätzen zur ‚Lesbarkeit der Welt‘ sollen deshalb im Folgenden einige Herausforderungen an eine digitale Literatur und Literaturwissenschaft formuliert werden. Literatur und Literaturwissenschaft, die sich unter den digitalen Bedingungen ohnehin annähern werden, sollen dabei unter den Prämissen einer partiellen Überlappung in den Blick genommen werden.⁸⁰

4 Herausforderungen einer digitalen Literatur(wissenschaft)

Es „ist in einer kurzen Zeitspanne, in jener, die uns von den siebziger Jahren trennt, ein neuer Mensch geboren worden. Er oder sie [...] kommuniziert nicht mehr auf die gleiche Weise, nimmt nicht mehr dieselbe Welt wahr, lebt nicht mehr in derselben Natur, nicht mehr im selben Raum“ – mit diesen Worten beschreibt der Philosoph Michel Serres die medialen und sozialen Kommunikationsformen einer seit etwa der Jahrtausendwende in Erscheinung tretenden „vernetzten Generation“.⁸¹ Welche Veränderungen die Lesbarkeit von Schrift und damit Literatur und Literaturwissenschaft unter solchen Bedingungen erfahren, sei in den folgenden fünf Punkten thesenhaft skizziert.

4.1 Netzwerke

Das von Lyotard inszenierte Experiment der *Épreuves d’écriture* führte die Beteiligten in einer digitalen Vernetzung zusammen, die zuallererst die Revision bereits verfasster Beiträge befördern sollte; die entsprechende Regel lautete: „il vous appartient [...] d’enchaîner à quelque fin que ce soit (réfuter, compléter, moduler, etc.), sur vos premières définitions“.⁸² Das so verstandene ‚Verknüpfen‘ („enchaîner“) veranlasste die Beteiligten, ihre Einträge zu überarbeiten, was

⁸⁰ Entsprechend verfährt ja auch Lyotard in seinem Nachwort, wenn er dieses den vorausgehenden Texten gleichstellt und einräumt: „le présent Post-scriptum n’est pas à lire comme un jugement, ni même un bilan, il faut le compter au nombre des textes qui le précèdent“ (Lyotard/Chaput [Anm. 14], 259; vgl. auch Abschn. 3, 39).

⁸¹ Michel Serres, *Petite Poucette*, Paris 2012; ins Dt. übers. von Stefan Lorenzer, *Erfindet euch neu! Eine Liebeserklärung an die vernetzte Generation*, Berlin 2013, hier: 15.

⁸² Lyotard/Chaput (Anm. 14), 7; vgl. auch Abschn. 3, 38.

sich gelegentlich in der Bereitstellung neuer Textversionen niederschlug. Ein solches Vorgehen ist etwa in einem lakonisch-frivolen Gedicht des Soziologen Jean-Claude Passeron belegt („on vit avec le vent virer structuraliste/Une marée, une manif [...]“), das der Verfasser, angeregt durch einen Beitrag des Philosophen François Recanati (über den fiktiven Fund eines Baudelaire-Gedichts), nachträglich in seiner Binnenstruktur erweitert hat.⁸³ Diese Interaktion erfolgt bezeichnenderweise unter dem Stichwort *métamorphose*, und sie wird von der Philosophin Christine Buci-Glucksmann mit Begriffen der Molekularbiologie (*mutants*) und des Schriftwesens (*écriture*) sowie mit einem hintersinnigen Neologismus (*difficulture*) kommentiert: „Serions-nous devenus des mutants? [...] Mais est-ce bien encore la dernière version? Métamorphose infinie [...], Écriture d'écritures [...], Difficulture“.⁸⁴

Dieses Beispiel illustriert nicht nur anschaulich die Interaktionsweise der an dem Schreibexperiment Beteiligten, sondern es belegt mit Passerons Gedichtbeitrag auch die Unfestigkeit der unter digitalen Bedingungen produzierten Texte. Lyotard verknüpfte diese Unfestigkeit konsequent mit dem Konzept eines multi-relationalen Textgefüges, das im Zusammenwirken der Beteiligten entsteht und zu vielfältigen, letztlich unplanbaren und im Prinzip ‚endlosen Metamorphosen‘ (vgl. das Zitat von Buci-Glucksmann) führen kann. Eine solche Netzwerkbildung erinnert an den Begriff des ‚Rhizoms‘ („Rhizome“) als eines unkontrolliert wuchernden Wurzelwerks, den Gilles Deleuze und Félix Guattari wenige Jahre zuvor eingeführt hatten.⁸⁵

Dieser rhizomartige Netzwerkbegriff hat eine syntaktische und eine pragmatische Dimension, da einzelne Zeichenfolgen und die darin erfolgenden Aussagen verknüpft werden (Syntagmatik), dies jedoch abhängig von der Interaktion der Zeichenbenutzer (Pragmatik). Bedeutung (Semantik) erwächst erst aus diesem multirelationalen Wechselspiel. In der Einleitung zu den *Épreuves d'écriture* erinnert Lyotard an die Großstadterfahrung, wie sie Walter Benjamin unter anderem im Rückgriff auf Baudelaire in seinem ‚Passagenwerk‘ festgehalten hat. Leitbegriffe wie „la rue“, „la foule“, „[le] nombre et [...] l'isolement“, „[le] passant“ kennzeichnen diese Erfahrung, die auch für den Verkehr und die Kommunikation in digitalen Netzwerken Geltung beanspruchen dürfen.⁸⁶ Entsprechend emphatisch fallen die Einträge zum Stichwort *réseau* („Netz“) in den

⁸³Ebd., 146–149 (mit den folgenden Autorensiglen und Datumsangaben: PASS. 122, 09 OCT.; RECA. 156, 28 NOV.; PASS. 130, 14 DEC.).

⁸⁴Ebd., 150 (BUCI. 088, 18 DEC.).

⁸⁵Vgl. Gilles Deleuze/Félix Guattari, *Rhizome. Introduction*, Paris 1976; wieder in: Dies., *Capitalisme et schizophrénie. Mille plateaux*, Paris 1980, 9–37; ins Dt. übers. von Dagmar Berger/Clemens-Carl Haerle/Helma Konyen, *Rhizom*, Berlin 1977.

⁸⁶Vgl. Lyotard/Chaput (Ann. 14), 6.

Épreuves d'écriture aus, wo das Netzwerk in seinen vermittelnden, distributiven, aber auch dissoziierenden Funktionen definiert wird.⁸⁷

In den Netzwerken führt die elektronische Verfügbarkeit von Texten zu einer enormen Beschleunigung des Datenaustauschs. Die Reproduktion des Gelesenen trägt den Anschein der zeitlichen, mitunter auch medialen ‚Unmittelbarkeit‘; aus den Lesern können auf diese Weise die im Nachwort evozierten Co-Autoren werden: „l’informatique autorise la reproduction immédiate“.⁸⁸ Es ist davon auszugehen, dass die Leerstellen, wie sie Iser als zentrales Konstituens des ‚Akts des Lesens‘ beschrieben hat, in den Sog dieser Beschleunigung geraten. Der digitale Kontext ermöglicht ein ‚unmittelbares‘ Füllen der Leerstellen durch Leser, die in der Rolle von Co-Autoren diese Füllung ihrerseits verschriftlichen. Dabei ist zu berücksichtigen, dass auch Iser's Konzept der Leerstellen die Idee eines (textuellen) Netzwerks zugrunde liegt, dessen Bestandteile durch Verknüpfungsleistungen miteinander in Bezug gesetzt werden. Erfolgt dieser Vorgang in einem digitalen Netzwerk (Syntagmatik), an dem multiple Benutzergruppen teilhaben (Pragmatik), so vervielfacht sich die Füllung der Leerstellen in einem potenziell ‚immediaten‘ Verschriftlichungsvorgang. Leerstellen werden dabei nicht nur ‚unmittelbar‘ gefüllt, sondern es werden zugleich immer neue Leerstellen produziert, die eine neuerliche Füllung ermöglichen. Die Doppelung des formulierten Textes durch Unformuliertes, wie sie Iser mit dem Begriff der ‚Negativität‘ beschreibt,⁸⁹ wird dabei gleichermaßen beseitigt wie jeweils neu reproduziert. Unter digitalen Bedingungen gilt diese ‚immediate‘ Füllung und Reproduktion für alle Leser, die ihrerseits zu Reproduzenten von Texten werden, also für nicht-professionelle und professionell-wissenschaftliche Leser gleichermaßen.

4.2 Simulation

Die im vorigen Abschnitt angesprochene Doppelung lässt sich (unter einer veränderten Verwendung dieses Begriffs) im digitalen Kontext auch für jene Form der ‚Erfahrung‘ veranschlagen, wie sie Blumenberg unter der Perspektive der ‚Lesbarkeit der Welt‘ in Betracht gezogen hat. Für das enzyklopädische Zeitalter des 18. Jahrhunderts setzte Blumenberg eine Verdoppelung der Welt bzw. der Natur in den Bänden der Enzyklopädie an; letztere stehen auf diese Weise „in Idealkonkurrenz mit dem ‚Buch der Natur““.⁹⁰ Die digitale Repräsentation der Welt hat eine solche

⁸⁷ Ebd., 194: „se brancher“ – „être branché“ (Philippe Curval, CURV. 077, 08 OCT.), „réseaux de distribution“ (Bruno Latour: LATO. 103, 09 OCT.), „le concept de la dissociation même“ (Jacques Derrida, DERR. 125, 10 OCT.).

⁸⁸ Ebd., 62, Philippe Curval unter dem Stichwort *écriture* (CURV. 104, 06 DEC.); zu den Co-Autoren das Nachwort ebd., 261, und Abschn. 3, 39 f.

⁸⁹ Vgl. Abschn. 2, 32.

⁹⁰ Vgl. Abschn. 2, 34.

Verdoppelung längst umgesetzt; sie bietet eine Universalbibliothek des Wissens, in der ein Maßstabsverhältnis von 1:1 gegenüber der sogenannten empirischen Wirklichkeit möglich scheint. Selbst eine Vervielfachung des Maßstabs auf der Zählerseite in Größenordnungen wie 10.000:1 oder 100.000:1 scheint potenziell denkbar – dies unter der Voraussetzung, dass die dabei erfasste empirische ‚Wirklichkeit‘ als eine simulierte kodiert und zur Darstellung gebracht wird.

In den *Épreuves d'écriture* wird der Weg von der empirischen Wirklichkeit zu deren gleichsam kartographischer Repräsentation in einem kurzen Dialog nachgezeichnet. Am Ende der Statements zum Stichwort *réseau* findet sich ein literarischer Beitrag des Schriftstellers Michel Butor, der eine Reise nach Nouméa in Neukaledonien evoziert. Aus kurzen Dialogpartien wird deutlich, dass zwei miteinander vertraute Personen eben einen Langstreckenflug um den halben Erdball antreten. Nach dem Start erscheinen den beiden Reisenden die bekannten Regionen Frankreichs mit den burgundischen Weinbergen und Flussläufen als eine wenn auch mit Unbestimmtheit wahrgenommene Landkarte: „Bientôt la Bourgogne. – Les vignes. – La Saône. – Tu es sûre que c'est la Saône ? – Je ne sais pas.“⁹¹ In diesem literarischen Experiment wird eine Erfahrung beschrieben, die in der Perspektive des Flugs aus der empirisch-geographischen Wirklichkeit herausführt und die Wahrnehmung einer Karte simuliert.

Unter dem Stichwort *simulation* werden die „machines simulant le vivant“ thematisiert.⁹² Viele der Beiträge sind dabei zeichentheoretisch orientiert. Analog zur Unterscheidung von ‚Bezeichnendem‘ und ‚Bezeichnetem‘ trennt der Philosoph Michel Tibon-Cornillot die Bereiche von Simulation und Simuliertem („le monde du simulateur, celui du simulé“); hinsichtlich der elektronischen Möglichkeiten des Computers („pour les machines“) unterscheidet er zwischen der künstlichen Darstellbarkeit („[le monde] de l'art, de la technique, bref de la culture“) und dem Darzustellenden („le vivant, la nature“).⁹³ Eine andere Sichtweise eröffnet der Naturwissenschaftler Paul Caro mit Erläuterungen zu der oben erwähnten Vergrößerung des Maßstabsverhältnisses: Die in einem elektronischen Mikroskop erfassten atomaren Strukturen müssen neu berechnet werden, um für das Auge überhaupt sichtbar zu werden. Die Instrumente passen sich den menschlichen Sinneskräften an und damit der mit den Sinnen wahrgenommenen Struktur einer Welt, in der – nach Descartes – ‚uns die Sinne täuschen‘.⁹⁴

Für die elektronische Darstellbarkeit der Welt hat dies eine Konzentration auf die bereits erwähnte Zeichenhaftigkeit der Simulation, genauer auf deren Bezeichnungsseite, zur Folge. Der Soziologe Bruno Latour diagnostiziert unter dem Stichwort *langage* im Zeitalter des Computers eine eigentliche Sprachrevolution:

⁹¹ Lyotard/Chaput (Anm. 14), 195 (BUTO 049, 30 NOV.).

⁹² Michel Tibon-Cornillot, ebd., 207 (TIBO. 184, 10 OCT.).

⁹³ Ebd.

⁹⁴ Vgl. ebd., 208: „Ainsi, nos instruments, fabriqués pour compenser la faiblesse de nos sens, doivent-ils être, eux aussi, adaptés à la structure réelle d'un monde dans lequel, comme le dit Descartes, ‚les sens nous trompent‘“ (CARO. 063, 28 NOV.).

Was einst dem Bereich des Materiellen, Realen, Referenziellen zugeordnet war, hat sich auf jenen des Codes, der Sprache, des Zeichens verschoben. Die Sprache als Bezeichnendes selbst aber wird technisiert: auf Mikrochips gespeichert, auf Bits beschränkt und von Elektronen gesteuert.⁹⁵ Die in der Digitalisierung ermöglichte Simulation von Wirklichkeit tendiert dabei dazu, sich auf den Code als Zeichenträger zu reduzieren, während ihr die Referenzialität abhandenkommt oder zum bloßen durch Zeichen produzierten Konstrukt gerät. Ein solcher Konstrukt-Charakter aber führt dazu, dass Literatur und Literaturwissenschaft in einen Kreislauf der Selbstbezüglichkeit geraten können.

4.3 Kombinatorik – Aleatorik

Die Möglichkeiten der Produktion und Reproduktion des Codes als Zeichenträger erscheinen unbegrenzt. Die dabei wirksame Kombinatorik ist prinzipiell schon in den Entwürfen aus dem Kontext der Gruppe *Poetik und Hermeneutik* angelegt, wenn dort von den nicht festgelegten Kombinationsverfahren im *Akt des Lesens* (Iser) und der durch Kombinatorik erzeugbaren Universalbibliothek als Dispositiv der *Lesbarkeit der Welt* (Blumenberg) die Rede ist.⁹⁶ Unter den Voraussetzungen digitaler Texterzeugung potenzieren sich diese Kombinationsmöglichkeiten, da die gebildeten Zeichensequenzen in ihrer Unfestigkeit fortlaufend reproduziert und reorganisiert werden können. Aus der Tatsache, dass dieser Prozess nicht bzw. nicht zwingend von einem konsistenten Organisator (einem Individuum oder einer kohärenten Gruppe) gesteuert wird, ergibt sich ein hohes Maß an Beliebigkeit, das zu unvorhersehbaren, zufälligen Ergebnissen führen kann.

In den *Épreuves d'écriture* findet sich dafür, wie übrigens schon bei Iser im Zusammenhang mit dem ‚Akt des Lesens‘,⁹⁷ der Begriff der ‚Aleatorik‘. Im Nachwort weist Lyotard explizit auf diesen spezifischen Aspekt des an den Bildschirmen durchgeführten kollektiven Schreibexperiments hin: „l'aléa y est inclus comme un risque à courir“.⁹⁸ Unter den behandelten Stichwörtern kehrt der Begriff mit einer gewissen Regelmäßigkeit wieder.⁹⁹ Im Kontext der vorausgehenden Ausführungen seien davon nur zwei Stellen herausgegriffen, die sich ihrerseits auf das Schreiben in einem Computernetzwerk beziehen: Unter dem Stichwort *simulation* verweist Christine Buci-Glucksmann auf eine in der Postmoderne wirksame Semiotik der Oberfläche („chimères sémiotiques“ – „le

⁹⁵Vgl. ebd., 111: „Pendant vingt ans, nous avons assisté, en Europe, à une révolution du langage. Tout ce qui était du côté des matières, des réalités, des référents, a basculé du côté du code, du langage, du signe. [...] Brusquement, le langage est devenu codé par des puces, référé à des bits, manipulé par des électrons“ (LATO. 111, 09 OCT.).

⁹⁶Vgl. Abschn. 2, 32 und 34.

⁹⁷Vgl. Abschn. 2, 32.

⁹⁸Lyotard/Chaput (Anm. 14), 262.

⁹⁹Vgl. ebd., 15, 38, 53, 101, 130, 181, 209 u. 262.

mystère est dans la surface“), die mit einer prinzipiellen Unabschließbarkeit und Unvollständigkeit („incomplétude“) einhergehe – dies aber impliziere die Hinwendung zum Unfesten und Aleatorischen: „[l]e retour à l’instable et à l’aléatoire“. ¹⁰⁰ Und im Zusammenhang mit dem Stichwort *auteur* hält der Ökonom Marc Guillaume fest, dass auch die Autorschaft einer solchen Aleatorik unterliege: „l’auteur peut devenir improbable, aléatoire“. ¹⁰¹

Der Begriff des Autors, der im französischen Kontext der Zeit durch die wegweisenden Beiträge von Roland Barthes (*The Death of the Author – La Mort de l’auteur*, 1967/68) und Michel Foucault (*Qu’est-ce qu’un auteur*) vorgeprägt ist, ¹⁰² erscheint im Kontext der *Épreuves d’écriture* unter dem Primat einer im digitalen Medium vollzogenen Diffundierung von Co-Autorschaften. ¹⁰³ Derrida bringt eben diesen Sachverhalt in einem einschlägigen Statement zum Ausdruck, wenn er im Kontext seiner Einlassung auf das Schreibexperiment fragt: „Y a-t-il un auteur dans cette entreprise commune?“ ¹⁰⁴ Das von Barthes und Foucault in den genannten Beiträgen diagnostizierte Verschwinden des Autors wird dabei konsequent mit den neuen technischen Möglichkeiten konfrontiert: „Ladite disparition de l’auteur passe peut-être toujours par l’expérience d’un tel dispositif socio-technique (machine à traitement de textes, central téléphonique anonyme, etc.)“. ¹⁰⁵ Zugleich macht Derrida darauf aufmerksam, dass der durch die technischen Möglichkeiten wenn nicht preisgegebene, so doch relativierte Begriff der Autorschaft im Kontrast zur persistenten Legalität des Autorenrechts steht: „tout reste encore signé, personne n’a le droit de toucher au texte de l’autre, notre copyright est très protégé comme au bon vieux temps de la modernité (17^e–20^e siècles)“. ¹⁰⁶

Die Erwähnung des Copyrights im Kontext der Autorschaft verweist auf ein besonderes Prinzip der Kombinatorik, das oben bereits mit dem Beitrag von Buci-Glucksmann zur „incomplétude“ digitaler Oberflächenphänomene (unter dem Stichwort *simulation*) angeklungen ist: Texte, aber auch Artefakte wie Bilder und Piktogramme, unterliegen einer fortlaufenden Reproduktion, innerhalb deren sie dupliziert, multipliziert und in der Multiplikation verändert werden können.

¹⁰⁰ Ebd., 209 (BUCI. 084, 11 DEC.).

¹⁰¹ Ebd., 15 (GUIL. 086, 09 OCT.).

¹⁰² Vgl. den zunächst in englischer Sprache erschienenen Beitrag von Roland Barthes, „The Death of the Author“, in: *Aspen Magazine* 5–6 (1967); französisch: „La Mort de l’auteur“, in: *Manteia* (1968), 12–17; Michel Foucault, „Qu’est-ce qu’un auteur“, in: *Bulletin de la Société française de Philosophie* 63 (1969), 73–104. Zum Begriff stellvertretend Fotis Jannidis/Gerhard Lauer/Matias Martinez u. a., „Autor und Interpretation“, in: Dies., *Texte zur Theorie der Autorschaft*, Stuttgart 2000, 7–34, mit Abdruck deutscher Übersetzungen der Beiträge von Barthes und Foucault, 181–197 u. 198–229.

¹⁰³ Vgl. Abschn. 3, 39 f., und 4.1, 43.

¹⁰⁴ Lyotard/Chaput (Anm. 14), 19, ebenfalls unter dem Stichwort *auteur* (DERR. 139, 05 DEC.).

¹⁰⁵ Ebd.

¹⁰⁶ Ebd.

Diese immer nur auf vorausliegende Zeichen verweisende Reproduzierbarkeit führt zu einer Umkehrung etablierter semiotischer Hierarchien („renversement des hiérarchies établies“), konkret zu einer Relativierung konventioneller Polaritäten der Kategorien von Bezeichnendem und Bezeichnetem: von Simulation und Realität, von Bild und Referent – „une relativisation des oppositions classiques entre simulation et réel, image et référent“.¹⁰⁷ Die elektronische Reproduzierbarkeit ermöglicht vielmehr die Kopie von Zeichenserien: „images d’images“, wie es Buci-Glucksmann am Beispiel ikonographischer Artefakte nennt.¹⁰⁸

Versteht man ‚Kopie‘ in ihrer etymologischen Bedeutung als ‚Menge‘ („copia“), so fällt auf, dass Michel Butor – im Bezug auf Texte – die Frage nach der zählbaren Menge von (Druck-)Exemplaren im digitalen Zeitalter als überholt ansieht: „(l)a question du nombre d’exemplaires serait définitivement dépassée“; der Text könne standortunabhängig überall zur selben Zeit sein: „le texte pourra être pratiquement partout à la fois“.¹⁰⁹ In einem Formexperiment spielen Paul Caro und Jean-Claude Passeron performativ die Möglichkeit durch, einen identischen Text, der bezeichnenderweise mit den Worten „(le) plaisir du *rewriting*“ endet, einmal unter dem Namen des einen, dann unter jenem des anderen Autors figurieren zu lassen.¹¹⁰ Der Informatiker Mario Borillo reflektiert in einem Beitrag, der in einem Anhang mit Texten steht, die außerhalb des experimentellen Computernetzwerks entstanden sind,¹¹¹ über die Eigenart elektronisch hergestellter Kopien: Deren Bandbreite bewege sich zwischen ununterscheidbarer Identität nach dem Muster eineiiger Zwillinge („des jumeaux monozygotes indiscernables“) und einer kontrollierten Variabilität („variations contrôlées“), die gerade die Hervorbringung identischer Exemplare ausschließe.¹¹²

Die Reproduzierbarkeit von Artefakten in identischen oder variablen Kopien stellt eine der zentralen Herausforderungen des digitalen Zeitalters dar. Sie bestimmt sowohl die Anfertigung von Texten als auch deren Weiterverarbeitung bis hin zu den Verfahren der (literaturwissenschaftlichen) Analyse und Interpretation. Letztere nehmen ihrerseits auf Prätexte als ihre Objekte oder als Vorlagen ihrer Argumentation Bezug. Die Rekombination von Zeichenfolgen in Form von Zitaten und Textmontagen bis hin zu aneignenden Formen wie Pastiche oder Plagiat spielt dabei eine wichtige Rolle. Eng damit verbunden sind Aspekte kollektiver bzw. fragmentierter Autorschaft und diffundierender Autorfunktionen. Reproduzierbarkeit und Autorschaft unterstehen in der digitalen Kultur den Gesetzen einer aleatorisch geprägten Rekombinatorik.

¹⁰⁷Ebd., 209 (BUCI. 084, 11 DEC.).

¹⁰⁸Ebd.

¹⁰⁹Ebd., 143, unter dem Stichwort *mémoire* (BUTO. 046, 30 NOV.).

¹¹⁰Ebd., 55, unter dem Stichwort *écriture* (einmal mit dem Signet CARO. 020, 08 OCT., dann als PASS. 115, 09 OCT.).

¹¹¹„Annexes“, ebd., 231–258.

¹¹²Ebd., 240, unter den Stichwörtern *écriturelangage* (BORI. 10 DEC.).

Die Kopie als Resultat einer kulturellen Praxis der Reproduktion findet dabei gerade in der aktuellen kultur- und literaturwissenschaftlichen Wahrnehmung eine hohe Aufmerksamkeit.¹¹³ Der Vorgang des Kopierens reicht freilich weit zurück in die Geschichte der Produktion und Reproduktion von Artefakten. Als ein einigermaßen willkürlich ausgewähltes Dokument sei ein Abschnitt aus Flauberts nachgelassenen Notizen zu dem unvollendeten Roman *Bouvard et Pécuchet* (1880) angeführt, in dem mit satirischem Unterton eine universelle Kopierpraxis evoziert wird, die zugleich jeden reflexiv-kritischen Anspruch preisgibt:

Pas de réflexion ! copions ! Il faut que la page s'emplisse, que le < monument > se complète. – Égalité de tout, du bien et du mal, du beau et du laid, de l'insignifiant et du caractéristique.¹¹⁴

4.4 Künstliche Intelligenz

Die in dem Flaubert-Zitat angesprochene Mechanik und Totalität des Kopierens geht im digitalen Zeitalter mit der Schaffung universeller elektronischer Archive einher, die das individuelle Gedächtnis nicht nur entlasten, sondern dessen Kapazitäten in einem gewaltigen Maß übersteigen. Michel Serres hat diese Auslagerung des Gedächtnisspeichers an elektronische Datenträger im Bild des frühchristlichen Märtyrers Dionysius beschrieben, der (gemäß der Legendensammlung *Legenda aurea*) seinen Kopf nach der Enthauptung auf dem Pariser Montmartre nordwärts bis zum Ort des später nach ihm benannten Heiligtums St-Denis getragen haben soll: „Wie der heilige Dionysius seinen halslosen Kopf, so hält sie [gemeint ist die von Michel Serres im Märchenmotiv des kleinen Däumlings imaginierte ‚vernetzte Generation‘] ihre vormals internen, nun externalisierten Fähigkeiten in Händen“.¹¹⁵ Die ausgelagerte Speicherung ermöglicht es, Daten jederzeit verfügbar zu halten, sie mit komplexen Suchmaschinen zu erschließen, und stellt damit traditionelle Techniken und Tugenden des Wissenserwerbs infrage. Gelangen wir damit auch an „das Ende der Ära des Wissens?“ – überlegt Serres lakonisch.¹¹⁶

In den *Épreuves d'écriture* begegnet anstelle des von Serres gewählten Bildes der ‚Enthauptung‘ der Begriff der ‚Prothese‘, dem ein eigenes Stichwort gewidmet ist. Paul Caro bezeichnet die Medien des elektronischen Zeitalters als „prothèses de nos sens et de nos organes moteurs“ und konstatiert, wiederum in biologischer

¹¹³Vgl. stellvertretend Gisela Fehrmann/Erika Linz/Eckhard Schumacher u. a. (Hg.), *Originalkopie. Praktiken des Sekundären*, Köln 2004; Hillel Schwartz, *The Culture of the Copy. Striking Likenesses, Unreasonable Facsimiles*, New York 2014 [1996]; Darren H. Hick/Reinhold Schmücker, *The Aesthetics and Ethics of Copying*, London u. a. 2016.

¹¹⁴Gustave Flaubert, *Bouvard et Pécuchet, avec les fragments du ‚second volume‘, dont le ‚Dictionnaire des idées reçues‘, chronologie, présentation, notes, dossier, bibliographie* par Stéphanie Dord-Crouslé, Paris 1999, 401.

¹¹⁵Michel Serres (Anm. 81), 27 f.

¹¹⁶Ebd., 29.

Begrifflichkeit, eine eigentliche Mutation in Bezug auf die memorialen Kräfte des Menschen: „une véritable mutation au sens biologique du terme“.¹¹⁷ Die digitale Auslagerung des Gedächtnisspeichers, so der Schriftsteller Philippe Curval unter dem Stichwort *mémoire*, sei ein Zeichen der Ermüdung und ein Vorgang, der die mit dem Gedächtnisverlust einhergehenden Beschwerden eher kaschiere: „Signe de fatigue. La mémoire est actuellement informatisée pour éviter les troubles physiologiques qu’engendre sa perte“.¹¹⁸

Neben ihrer das Gedächtnis entlastenden Funktion ermöglichen die auf elektronischen Datenträgern abgelegten Texte, Bilder, Artefakte auch eine neue Art des Lesens und der Wahrnehmung. Die gespeicherten Daten werden ihrerseits mit elektronischen Verfahren weiterverarbeitet. Im Hinblick auf Texte kann eine solche Weiterverarbeitung nur noch bedingt als ‚Lesen‘ bezeichnet werden, eher als ein ‚Sammeln‘ („jegere“ in seiner ursprünglichen Bedeutung).¹¹⁹ Texte werden in der Verarbeitung durch elektronische Werkzeuge weniger ‚gelesen‘ als ‚benutzt‘; sie werden kopiert, verknüpft, fragmentiert.¹²⁰ Wenn Blumenberg im Blick auf die Leibniz’sche *Ars combinatoria* unterstellt, dass eine ideale „kombinatorische Bibliothek [...] keinen Leser“ mehr hätte,¹²¹ so scheint diese Vision in der digitalen Kombinatorik umgesetzt. An die Stelle des konventionellen Lesers und seiner kognitiven Fähigkeiten treten Verarbeitungspraktiken einer Künstlichen Intelligenz. Die mit dem Namen des Literaturwissenschaftlers Franco Moretti verbundene Methode des *Distant Reading* beruht maßgeblich auf solchen technisierten Praktiken der Lektüre.¹²²

Welche enormen kreativen Potenziale der Künstlichen Intelligenz eignen, zeigte sich in dem 1997 als Medienereignis inszenierten Sieg des Schachcomputers *Deep Blue* über den Schachweltmeister Garri Kasparov.¹²³ Im Bereich der Textproduktion gibt es visionäre Ansätze einer aleatorischen Texterzeugung, die bis in die Zeit der *Ars combinatoria* des 17. Jahrhunderts zurückreichen. Blumenberg erinnert in diesem Zusammenhang an das Spottgedicht des sächsischen Schriftstellers Daniel Wilhelm Triller (1695–1782) über einen Lettern setzenden Affen,

¹¹⁷Lyotard/Chaput (Anm. 14), 191 (CARO. 035, 08 OCT.).

¹¹⁸Ebd., 139 (CURV. 076, 08 OCT.).

¹¹⁹Vgl. Abschn. 1, 28.

¹²⁰Vgl. dazu den oben (Anm. 4) erwähnten Beitrag von Braun u. a. im vorliegenden Band, bes. Abschn. 1.3, 290, und 2, 295.

¹²¹Blumenberg (Anm. 11), 133; vgl. Abschn. 2, 33 f.

¹²²Vgl. Moretti (Anm. 5), und Abschn. 1, 28.

¹²³Vgl. dazu zuletzt Martina Heßler, „Der Erfolg der ‚Dummheit‘. Deep Blues Sieg über den Schachweltmeister Garri Kasparov und der Streit über seine Bedeutung für die Künstliche Intelligenz-Forschung“, in: *NTM. Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin* 25/1 (2017), 1–33; Manuela Lenzen, *Künstliche Intelligenz. Was sie kann & was uns erwartet*, München 2018, 107 f.

der „unter viel Wertlosem auch etwas Lesbares zustande zu bringen“ meint.¹²⁴ In dem bereits erwähnten Anhang zu den *Épreuves d'écriture* fragt der Informatiker Mario Borillo nach den Möglichkeiten einer „machine-auteur“.¹²⁵ Deren Leistungsfähigkeit müsste sich am Maßstab der ‚Originalität‘ (französisch „*inédit*“, im Sinne von: ‚noch nie da gewesen‘) bemessen: „à l'aune le la production d'un texte, d'une œuvre littéralement inédits“ – einer ‚Originalität‘ freilich, die ihrerseits ihr kombinatorisches Potenzial offenbart („la manifestation de l'extraordinaire fécondité combinatoire“).¹²⁶ In einem Projekt mit dem Titel *Atelier de Littérature Assistée par Mathématique et Ordinateur (ALAMO)* haben Borillo und sein Team die Möglichkeiten einer solchen künstlich erzeugten Literatur erprobt. Die Methoden und Resultate sind auf einer Website und in der Datenbank der *Bibliothèque nationale de France* dokumentiert.¹²⁷

Die aktuelle Diskussion zur Künstlichen Intelligenz ist über den in Borillos Statement greifbaren Enthusiasmus eines Originalitätsanspruchs technischer Kreativität hinausgegangen. Als „entscheidende Aufgabe einer geistes- oder kulturwissenschaftlichen Forschung“ zeichnet sich inzwischen ab, dass „die Unterschiede maschinellen und menschlichen Denkens genau zu beschreiben“ sind.¹²⁸ Es gilt, die Wirksamkeit menschlichen Denkens einerseits und die kalkulatorische Systematik des Computers andererseits in ihrer jeweiligen Eigenart zu erfassen und auf diese Weise in Interaktion zu bringen. Ob dabei die Produktion von Texten – abgesehen von ableitbaren Formen wie Übersetzungen – je zu einer mit der menschlichen Kreativität vergleichbaren ‚Originalität‘ gelangen wird, muss beim derzeitigen Stand offenbleiben. Hinsichtlich der Analyse literarischer Verfahren öffnet sich jedoch mit der sogenannten Künstlichen Intelligenz ein weites Feld, das vielversprechende Ergebnisse etwa im Hinblick auf die Erforschung von Schreibweisen und Textzuordnungen (z. B. zu ‚Schreibstilen‘, ‚Autoren‘, ‚Autorenkonfigurationen‘) erwarten lässt.

4.5 Die breite Zuhandenheit von Wissensvorräten

Mit Blick auf die Herausforderungen, vor welche Literatur und Literaturwissenschaft angesichts einer digitalen Textualität gestellt sind, ist abschließend auf ein Phänomen einzugehen, das als ‚breite Zuhandenheit von Wissensvorräten‘ bezeichnet werden soll. Der Begriff orientiert sich an dem Konzept der ‚breiten

¹²⁴Blumenberg (Anm. 11), 131, zu Daniel W. Triller, „Der Affe, ein seltsamer Buchdrucker, und ein Eremit“, in: Ders., *Neue Aesopische Fabeln*, Hamburg 1740, 237–239.

¹²⁵Liotard/Chaput (Anm. 14), 240 (BORI. 10 DEC.).

¹²⁶Ebd.

¹²⁷Vgl. <http://www.alamo.free.fr/> und http://data.bnf.fr/16531511/atelier_de_litterature_assistee_par_la_mathematique_et_les_ordinateurs/ (letzter Aufruf der Links 14.6.2019).

¹²⁸Heßler (Anm. 123), 27.

Gegenwart‘, das der Literaturwissenschaftler Hans Ulrich Gumbrecht seit der Jahrtausendwende propagiert und in verschiedenen Abhandlungen erläutert hat.¹²⁹ Das Konzept der ‚breiten Gegenwart‘ beinhaltet terminologisch eine räumliche und eine zeitliche Komponente; es bezieht sich auf die Erfahrung einer von Vergangenheiten überfluteten Gegenwart, die sich zugleich in zahlreichen Simultaneitäten ‚verbreitert‘.¹³⁰ Globalisierung und Technisierung befördern eine solche Erfahrung, wobei Gumbrecht die dabei im Spiel befindliche „Perfektion elektronischer Gedächtnisleistungen“¹³¹ äußerst kritisch sieht. Die in der antiken Rhetorik dem Redner zur Verfügung stehende „copia“ [als] Besitz eines ausgedehnten Wissensschatzes“ ist heute mit dem Computer „zu jenem Dispositiv geworden, das Wissen in früher unvorstellbarer Breite und Dichte abrufbar macht“.¹³² Dieses Phänomen soll hier in Anlehnung an Heideggers Begriff als ‚Zuhandenheit‘ (im Sinne des der menschlichen Verfügbarkeit „zunächst begegnenden innerweltlichen Seienden“) bezeichnet werden.¹³³ Die ‚breite Zuhandenheit von Wissensvorräten‘ bezieht sich mithin auf eine allgegenwärtige Verfügbarkeit von Wissen in den elektronischen Medien.

In den *Épreuves d'écriture* wird dieser Sachverhalt unter dem Stichwort *temps* in Ansätzen reflektiert. Borillo beobachtet, dass sich am Computer verschiedene Zeiteinheiten bündeln: jene des in den Daten verhandelten Gegenstands, jene der die Daten formalisierenden Strukturen (der Algorithmen und der zur Anwendung kommenden Programme) sowie jene der psycho-biologischen Befindlichkeit der Benutzer. Borillo folgert daraus, dass die Informatik einer vereinheitlichten Zeittheorie (vielleicht auch der Theorie einer vereinheitlichten Zeit?) bedarf: „L'informatique pose le problème d'une théorie unifiée de temps“.¹³⁴ Der Musikologe und Philosoph Daniel Charles seinerseits hält, allerdings ohne expliziten Bezug auf die Computertechnik, fest, dass die Zeit der Postmoderne nicht linear, nicht zielgerichtet und damit gewissermaßen in einem Stillstand der ‚Gleichzeitigkeit‘ befindlich sei: „Le temps post-moderne est non linéaire et non fléché. [...] Donc temps stable, équitemporel“.¹³⁵ Solche Diagnosen kommen dem Konzept der ‚breiten Gegenwart‘ und dem oben vorgeschlagenen Begriff der ‚breiten Zuhandenheit von Wissensvorräten‘ recht nahe.

¹²⁹ Vgl. stellvertretend Hans U. Gumbrecht, „Die Gegenwart wird (immer) breiter“, in: *Merkur* 55/629 (2001), 769–784; Ders., *Unsere breite Gegenwart*, aus dem Engl. übers. von Frank Born, Berlin 2010.

¹³⁰ Vgl. z. B. ebd., 16 u. 104 f. (hier mit Bezug auf Lyotards *La Condition postmoderne* [Anm. 60]).

¹³¹ Ebd., 16.

¹³² Ebd., 69.

¹³³ Vgl. Martin Heidegger, *Sein und Zeit*, Tübingen ¹⁵1993 [1927], 69 u. 88 (Zitat).

¹³⁴ Lyotard/Chaput (Anm. 14), 220 (BORI 065, 05 OCT.).

¹³⁵ Ebd., 221 (CHAR 079, 24 OCT.).

Die in den Statements der Beiträger zu den *Épreuves d'écriture* thematisierte ‚Vereinheitlichung‘ bzw. ‚Gleichzeitigkeit‘ des Temporalen hat auch Konsequenzen für das Verhältnis von literarischer Erfahrung und Erwartungshorizont, das Jauß zur Grundlage seines Rezeptionsästhetischen Ansatzes gemacht hat. Die ‚breite Zuhandenheit von Wissensvorräten‘ erschwert zumindest konzeptionell die diachrone Schichtung von historisch zu rekonstruierenden Erwartungshorizonten. Vielmehr dürfte für gegenwärtige Leser am Bildschirm die Synchronie des elektronisch Zuhandenen in hohem Maße dominant werden. Die einmal in das Netz von Wissensvorräten eingespeisten Informationen sind in ihrer Fülle (*copia*) alle ‚gleichzeitig‘ (*equitemporal*) verfügbar. Der Umgang mit ihnen bedarf eines diachronen Sonderwissens. Ebenso begünstigt das digitale Medium eine sofortige schriftgebundene Reaktion auf das in den elektronischen Netzwerken Erlesene. ‚Literarische Erfahrung‘ mündet dann tendenziell immer schon in ‚unmittelbare‘ (immediate)¹³⁶ Schreibakte; aus der Rezeption wird permanente Produktion. So deuten es Lyotard und seine Mitarbeiterinnen in ihrem Nachwort an („on place déjà le texte dans le réseau de sa circulation“);¹³⁷ und Bruno Latour bringt es mit dem Satz auf den Punkt: „[T]out le monde dévient scribe“.¹³⁸ Dies aber hat zur Folge, dass sich die Erwartungshorizonte der Lesenden unablässig verschieben, auch, dass sie in einem potenziell globalen Netzwerk, an dem alle Beteiligten gleichzeitig in einer ‚breiten‘ Streuung partizipieren, diffus und unübersichtlich werden.

Es bleibt abschließend zu fragen, wie es unter den Bedingungen einer ‚breiten Zuhandenheit von Wissensvorräten‘ um die Metapher der ‚Lesbarkeit der Welt‘ bestellt ist. Die auf elektronischen Trägern gespeicherten und an Bildschirmen abrufbaren Daten simulieren mit ihrer „illusion calligraphique“¹³⁹ eine Welt, deren Referenzialität zu entgleiten droht. Als eine simulierte unterliegt diese in elektronische Zeichensysteme gebrachte Welt stets schon einem metaphorischen ‚Als-ob‘. Die digital gebannte Welt ist ‚wie‘ ein Text zu lesen, oder sie ist – in metaphorischer Verkürzung – ein zu lesender Text. Was geschieht, wenn diese Verkürzung nicht mehr als metaphorische wahrgenommen wird, wenn sich – wie dies Blumenberg als Möglichkeit an der Problematik des genetischen Codes entwickelt hat – die „Lücke zwischen Metapher und Modell zu schließen“¹⁴⁰ scheint? Dann wäre die in den Computern simulierte Welt keine lesbare mehr, sondern sie würde die Illusion gewähren, eine in ihrer ‚Zuhandenheit‘ handhabbare Welt zu sein.

¹³⁶Vgl. Abschn. 4.1, 43.

¹³⁷Lyotard/Chaput (Anm. 14), 260; vgl. Abschn. 3, 39.

¹³⁸Ebd., 55, unter dem Stichwort *Écriture* (LATO. 106, 09 OCT.).

¹³⁹Philippe Curval unter den Stichwörtern *Dématérialisation/Métamorphose* ebd., 47 (CURV. 102, 31 OCT.).

¹⁴⁰Blumenberg (Anm. 11), 376; vgl. Abschn. 2, 35.

5 Ausblick

Als Herausforderung der Literatur und Literaturwissenschaft konkretisieren sich damit fünf Themenfelder, die vielfältig miteinander in Zusammenhang stehen und hier abschließend nochmals thesenartig benannt seien: Netzwerke bedingen eine neue Form der Interaktion im literarischen Prozess und damit neue Kategorien von Lesenden und Schreibenden. Die über die digitale Technologie erwirkte Simulation von Welt befördert die Dominanz des Codes als Zeichenträger – dies mit Konsequenzen sowohl für die Literatur als semiotisches System als auch für deren Analyse mit semiologischen Methoden. Kombinatorik und Aleatorik bewirken neue Konzeptionen der Autorschaft und der literarischen Produktion; dabei treten Formen der Reproduktion und der (digitalen) Kopie in den Vordergrund. Künstliche Intelligenz, die Auslagerung der Datenspeicherung an Maschinen, verändert die Analyse, aber auch die Produktion literarischer Daten. Die ‚breite Zuhandenheit von Wissensvorräten‘ führt zu einer Simultaneität literarischer Erfahrung, die Erwartungshorizonte nivelliert und zugleich in eine diffuse Sphäre des ‚Als-ob‘ verschiebt.

Die Erarbeitung dieser Themenfelder erfolgte im vorliegenden Beitrag in Auseinandersetzung mit einem digitalen Schreibexperiment, das mit seinem gut dreißigjährigen Abstand historischen Stellenwert besitzt. Der Blick auf die von Jean-François Lyotard 1984 inszenierten *Épreuves d'écriture* ermöglichte jedoch im Horizont der Gruppe *Poetik und Hermeneutik* Bezugnahmen auf literaturwissenschaftliche Paradigmen, wie sie das Verständnis einer ‚Lesbarkeit der Welt‘ vor der digitalen Wende bestimmten. Auf diese Weise sollte die Abgrenzung von konventionellen Konzepten des Lesens der prä-digitalen Ära profiliert werden. In der damit erzielten zeitlichen Distanz besteht die Chance, Tendenzen von Literatur und Literaturwissenschaft unter digitalen Bedingungen womöglich konkreter zu fassen, als dies mit einer Bestandsaufnahme jüngster Entwicklungen der Digitalität möglich ist. In einem nächsten Schritt wäre freilich genau dies zu leisten: nämlich die Beantwortung der Frage, welchen Einfluss die immer komplexer werdenden Standards der Speicherung, des Austauschs und der Erschließung gewaltiger Datenmengen auf das Selbstverständnis von Literatur und Literaturwissenschaft ausüben. Letztere sind durchaus im wörtlichen Sinne als eine Kombinatorik von *litterae* bzw. *lèttres* zu verstehen, die es im Lesen zu sammeln (*legere*) gilt. In diesem etymologischen Verständnis zeigen sich Chancen und Grenzen einer Literatur(wissenschaft), die sich auf das Zusammenwirken humaner und technisierter Praktiken bei der Produktion und Rezeption von Texten einlässt.

Literatur

Amslinger, Julia, *Eine neue Form von Akademie. ‚Poetik und Hermeneutik‘ – die Anfänge*, Paderborn 2017.

Aust, Hugo, „Lesen“, in: Harald Fricke/Georg Braungart/Klaus Grubmüller u. a. (Hg.), *Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*, Bd. 2: H–O, Berlin/New York 2000, 406–410.

- Bäcker, Iris, *Der Akt des Lesens – neu gelesen. Zur Bestimmung des Wirkungspotentials von Literatur*, Paderborn 2014.
- Barthes, Roland, „The Death of the Author“, in: *Aspen Magazine* 5–6 (1967); französisch: „La Mort de l’auteur“, in: *Manteia* (1968), 12–17.
- Benjamin, Walter, „Literaturwissenschaft und Literaturgeschichte“, in: Ders., *Ausgewählte Schriften*, Bd. 2: *Angelus Novus*, Frankfurt a. M. 1966 [1931], 450–456.
- Blumenberg, Hans, *Die Arbeit am Mythos*, Frankfurt a. M. 2001 [1979].
- Blumenberg, Hans, *Die Lesbarkeit der Welt*, Frankfurt a. M. 21983 [1981].
- Boden, Petra/Zill, Rüdiger (Hg.), *Poetik und Hermeneutik im Rückblick. Interviews mit Beteiligten*, Paderborn 2017.
- Campe, Rüdiger, „Lecture“, in: Harald Fricke/Georg Braungart/Klaus Grubmüller u. a. (Hg.), *Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*, Bd. 2: *H–O*, Berlin/New York 2000, 385–388.
- Deleuze, Gilles/Guattari, Félix, *Rhizome. Introduction*, Paris 1976; wieder in: Dies., *Capitalisme et schizophrénie. Mille plateaux*, Paris 1980, 9–37; ins Dt. übers. von Dagmar Berger/Clemens-Carl Haerle/Helma Konyen, *Rhizom*, Berlin 1977.
- Diderot, Denis/d’Alembert [d.i. Jean-Baptiste le Rond], *Encyclopédie, ou Dictionnaire raisonné des Sciences, des Arts et des Métiers*, Paris 1751–1780.
- Ette, Ottmar, *Der Fall Jaus. Wege des Verstehens in eine Zukunft der Philologie*, Berlin 2016.
- Fehrmann, Gisela/Linz, Erika/Schumacher, Eckhard u. a. (Hg.), *Originalkopie. Praktiken des Sekundären*, Köln 2004.
- Flaubert, Gustave, *Bouvard et Pécuchet, avec les fragments du ‚second volume‘, dont le ‚Dictionnaire des idées reçues‘, chronologie, présentation, notes, dossier*, bibliographie par Stéphanie Dord-Crouslé, Paris 1999.
- Foucault, Michel, „Qu’est-ce qu’un auteur“, in: *Bulletin de la Société française de Philosophie* 63 (1969), 73–104.
- Fricke, Harald/Braungart, Georg/Grubmüller, Klaus u. a. (Hg.), *Reallexikon der deutschen Literaturwissenschaft*, Neubearbeitung des Reallexikons der deutschen Literaturgeschichte, Berlin/New York 1997–2003.
- Gumbrecht, Hans U., „Die Gegenwart wird (immer) breiter“, in: *Merkur* 55/629 (2001), 769–784.
- Gumbrecht, Hans U., *Unsere breite Gegenwart*, aus dem Engl. übers. von Frank Born, Berlin 2010.
- Heidegger, Martin, *Sein und Zeit*, Tübingen ¹⁵1993 [1927].
- Heßler, Martina, „Der Erfolg der ‚Dummheit‘. Deep Blues Sieg über den Schachweltmeister Garri Kasparov und der Streit über seine Bedeutung für die Künstliche Intelligenz-Forschung“, in: *NTM. Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin* 25/1 (2017), 1–33.
- Hick, Darren H./Schmücker, Reinhold, *The Aesthetics and Ethics of Copying*, London u. a. 2016.
- Hui, Yuk/Broeckmann, Andreas (Hg.), *30 Years after ‚Les Immatériaux‘. Art, Science, and Theory* [Lüneburg 2015].
- Iser, Wolfgang, *Der implizite Leser. Kommunikationsformen des Romans von Bunyan bis Beckett*, München 1972.
- Iser, Wolfgang, *Der Akt des Lesens. Theorie ästhetischer Wirkung*, München 1976.
- Jakobson, Roman, „Closing Statement. Linguistics and Poetics“, in: Thomas A. Sebeok (Hg.), *Style in Language*, Cambridge, MA, 1960, 350–377.
- Jannidis, Fotis/Lauer, Gerhard/Martinez Matias u. a., „Autor und Interpretation“, in: Dies., *Texte zur Theorie der Autorschaft*, Stuttgart 2000, 7–34.
- Jannidis, Fotis/Kohle, Hubertus/Rehbein, Malte (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart 2017.
- Jauß, Hans R., *Literaturgeschichte als Provokation der Literaturwissenschaft*, Konstanz 1967 (*Konstanzer Universitätsreden* 3); wieder abgedruckt in: Ders., *Literaturgeschichte als Provokation*, Frankfurt a. M. 1970, 144–207.

- Lasswell, Harold D., „The Structure and Function of Communication in Society“, in: Lyman Bryson (Hg.), *The Communication of Ideas. A Series of Addresses*, New York 1948, 37–51.
- Lenzen, Manuela, *Künstliche Intelligenz. Was sie kann & was uns erwartet*, München 2018.
- Lyotard, Jean-François, *La Condition postmoderne. Rapport sur le savoir*, Paris 1979.
- Lyotard, Jean-François, *Le différend*, Paris 1983.
- Lyotard, Jean-François (Hg.), *Les Immatériaux. Album et inventaire*, Paris 1985.
- Lyotard, Jean-François/Chaput, Thierry (Hg.), *Épreuves d'écriture*, Paris 1985.
- Lyotard, Jean-François/Derrida, Jacques/Burkhardt, François, *Immaterialität und Postmoderne*, aus dem Franz. übers. von Marianne Karbe, Berlin 1985.
- Meyer, Thomas, „Lesbarkeit“, in: Robert Buch/Daniel Weidner (Hg.), *Blumenberg lesen. Ein Glossar*, Berlin 2014, 171–184.
- Moretti, Franco, *Distant Reading*, London 2013; ins Dt. übers. von Christine Pries, *Distant Reading*, Konstanz 2016.
- Müller, Jan-Dirk (Hg.), *„Aufführung“ und „Schrift“ in Mittelalter und Früher Neuzeit*, Stuttgart/Weimar 1996.
- Rautenberg, Ursula/Schneider, Ute, *Lesen. Ein interdisziplinäres Handbuch*, Berlin/Boston 2015.
- Sankt Viktor, Hugo von, „Eruditionis Didascalicae Libri Septem“, in: Migne, *Patrologia Latina* 176 (1854), 739–838.
- Schwartz, Hillel, *The Culture of the Copy. Striking Likenesses, Unreasonable Facsimiles*, New York 2014 [1996].
- Serres, Michel, *Petite Poucette*, Paris 2012; ins Dt. übers. von Stefan Lorenzer, *Erfindet euch neu! Eine Liebeserklärung an die vernetzte Generation*, Berlin 2013.
- Stalder, Felix, *Kultur der Digitalität*, Berlin 2016.
- Triller, Daniel W., *Neue Aesopische Fabeln*, Hamburg 1740.
- Warning, Rainer (Hg.), *Rezeptionsästhetik*, München 1975.
- Weimar, Klaus, „Literatur“, in: Harald Fricke/Georg Braungart/Klaus Grubmüller u. a. (Hg.), *Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*, Bd. 2: H–O, Berlin/New York 2000, 443–448.
- Weinrich, Harald, „Für eine Literaturgeschichte des Lesers“, in: *Merkur* 21/236 (1967), 1026–1038.
- Westemeier, Jens, *Hans Robert Jauf. Jugend, Krieg und Internierung*, Konstanz 2016.
- Winter, Thomas N., „Roberto Busa, S. J., and the Invention of the Machine-Generated Concordance“, in: *The Classical Bulletin* 75/1 (1999), 3–21.
- Wunderlich, Antonia, *Der Philosoph im Museum. Die Ausstellung ‚Les Immatériaux‘ von Jean François Lyotard*, Bielefeld 2008.
- Zumthor, Paul, *La Lettre et la voix. De la ‚littérature‘ médiévale*, Paris 1987.

Online-Ressourcen

Sämtliche digitalen Referenzen wurden letztmalig am 14.6.2019 eingesehen.

Atelier de Littérature Assistée par Mathématique et Ordinateur (ALAMO), <http://www.alamo.free.fr/> und http://data.bnf.fr/16531511/atelier_de_litterature_assistee_par_la_mathematique_et_les_ordinateurs/.

Digitalität in den Geisteswissenschaften, <http://digitalitaet-geisteswissenschaften.de/>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Empirie – Beschreibung – Interpretation. Über den Platz von Computermodellen in den hermeneutisch-historisch orientierten Literaturwissenschaften

Jonas Kuhn

Im Zuge der Positionsbestimmung und Entfaltung der *Digitalen Literaturwissenschaft* wird – jenseits der Unterstützung konventioneller Arbeitspraktiken mit digitalen Hilfsmitteln – über statistische Verfahren diskutiert und mit Werkzeugen experimentiert, die literaturwissenschaftlich relevante Texteigenschaften in einem größeren Untersuchungskorpus erfassen, beispielsweise zu Autorstil und Gattungsmerkmalen.¹

Mit sprach- und textanalytischen Werkzeugen werden wichtige Inhaltselemente wie Figurenerwähnungen² oder Figurenrede³ automatisch extrahiert, um etwa ein sogenanntes *Distant Reading* zu ermöglichen: Beispielsweise können

¹Justine Kao/Daniel Jurafsky, „A Computational Analysis of Style, Affect, and Imagery in Contemporary Poetry“, in: Proceedings of the Workshop on Computational Linguistics for Literature (Conference of the North American Chapter of the Association for Computational Linguistics. Human Language Technologies, NAACL-HLT), Montréal 2012, 8–17; Fotis Jannidis/Gerhard Lauer, „Burrows’s Delta and Its Use in German Literary History“, in: Matt Erlin/Lynne Tatlock (Hg.), *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*, Rochester 2014, 29–54.

²Hardik Vala/David Jurgens/Andrew Piper u. a., „Mr. Bennet, his coachman, and the Archbishop walk into a bar but only one of them gets recognized. On the difficulty of detecting characters in literary texts“, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, hg. von Association for Computational Linguistics, Lissabon 2015; Fotis Jannidis/Markus Krug/Isabella Reger u. a., *Automatische Erkennung von Figuren in deutschsprachigen Romanen. Digital Humanities im deutschsprachigen Raum (DHd)*, Graz 2015.

³David K. Elson/Kathleen R. McKeown, „Automatic attribution of quoted speech in literary narrative“, in: Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI’10) 2010, AAAI Press, 1013–1019.

J. Kuhn (✉)
Universität Stuttgart, Stuttgart, Deutschland
E-Mail: jonas.kuhn@ims.uni-stuttgart.de

Texteigenschaften durch Netzwerkanalysen⁴ oder einen Vergleich der Figuren-Charakterisierung innerhalb der Figurenrede⁵ über größere Textmengen aggregiert werden. Häufig werden hierfür korpusbasierte Computermodelle eingesetzt, die mithilfe von maschinellen Lernverfahren für die Vorhersage von Textinstanzen ‚trainiert‘ werden, auf die z. B. eine (operationalisierte) narratologische Analyse-kategorie zutrifft. Es gibt eine wachsende Zahl von Beispielen für komputationell anspruchsvolle Modellierungen, die auf großen Korpora von literarischen Texten entwickelt und angewendet wurden – wie beispielsweise die Modelle von David Bamman und Kollegen zu Figurentypen in englischsprachigen Romanen, für die statistischen Modelle auf über 15.000 Romanen trainiert wurden.⁶ Innerhalb der breiteren literaturwissenschaftlichen Fachdiskussion scheinen solche Ansätze jedoch bislang (noch?) keine zentrale Rolle zu spielen.

Es mangelt nicht an neueren Diskussionsbeiträgen zum vielschichtigen Spannungsverhältnis zwischen der *Digitalen Literaturwissenschaft* (bzw. *Digital Humanities*) und der etablierten Literaturwissenschaft,⁷ in dem sich unterschiedliche Erklärungen dafür ausmachen lassen, weshalb die komputationellen Methoden und Werkzeuge nicht zu einem rasanten Umdenken in großen Teilen des Gesamtfeldes führen – starkes Gewicht hat hier neben wissenschaftssoziologischen Gegebenheiten immer wieder die Abwägung zwischen Aufwand (Einarbeitung in die Methodik und Operationalisierung relevanter Analyse-kategorien) und dem zu erwartenden – nicht-trivialen – Ertrag;⁸ im Einzelfall mag diese Abwägung häufig berechtigterweise gegen den Einsatz von Computermodellen sprechen. Das Argument mit der vielleicht breitesten Akzeptanz für

⁴David K. Elson/Nicholas Dames/Kathleen R. McKeown, „Extracting social networks from literary fiction“, in: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, ACL'10, Stroudsburg, PA, 2010, 138–147; Franco Moretti, „Network Theory, Plot Analysis“, in: Stanford Literary Lab Pamphlets 2 (Mai 2011), <http://litlab.stanford.edu/LiteraryLabPamphlet2.pdf> (letzter Aufruf 22.09.2018); Peer Trilcke, „Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft“, in: Philip Ajouri/Katja Mellmann/Christoph Rau (Hg.), *Empirie in der Literaturwissenschaft*, Münster 2013, 201–247.

⁵Julian Brooke/Adam Hammond/Graeme Hirst, „Using Models of Lexical Style to Quantify Free Indirect Discourse in Modernist Fiction“, in: *Digital Scholarship in the Humanities*, 2016, <https://doi.org/10.1093/lc/fqw072>.

⁶U. a. in David Bamman/Ted Underwood/Noah A. Smith, „A Bayesian Mixed Effects Model of Literary Character“, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, MD, 2014, 370–379.

⁷Bspw. Trilcke (Anm. 4); Thomas Weitin, „Digitale Literaturwissenschaft“, in: *Deutsche Vierteljahresschrift für Literaturwissenschaft und Geistesgeschichte* 89/4 (2015), Sonderheft: Zur Lage der Literaturwissenschaft. Aktuelle Bestandsaufnahmen und Perspektiven, 651–656; Friedrich M. Dimpel, „Der Computerphilologe als Interpret – ein Teilzeit-Empiriker?“, Jan Borkowski/Stefan Descher/Felicitas Ferder u. a., *Literatur interpretieren. Interdisziplinäre Beiträge zur Theorie und Praxis*, Münster 2015, 339–359; Marcus Willand, „Hermeneutische Interpretation und digitale Analyse. Eine Verhältnisbestimmung“, in: Luisa Banki/Michael Scheffel (Hg.), *Lektüren. Positionen zeitgenössischer Philologie*, Trier 2017, 77–100.

⁸Trilcke führt dies für die literaturwissenschaftliche Netzwerkanalyse aus (Anm. 4), 203–208.

Computermodelle liegt auf der anderen Seite in deren Reliabilität, die es erlaubt, für beliebig große Textsammlungen strikt vergleichbare Analyseergebnisse zu erzeugen. So können Frequenzmuster bis in subtile Details erfasst werden, welche für menschliche Leser nicht unverzerrt (*reliably* oder ‚reliabel‘) wahrnehmbar sind, da deren Aufmerksamkeit durch inhaltliche Aspekte überlagert wird – die Erfolgsgeschichte stilometrischer Verfahren unterstreicht dies. Dank ihrer Reliabilität können Computeranalysen darüber hinaus auch der Problematik von Bestätigungsfehlern (dem *Confirmation Bias*) entgegengestellt werden, also der kognitionspsychologisch erklärbaren Tendenz, bei einer manuellen Textanalyse verstärkt diejenigen Eigenschaften zu sehen, die die eigene Hypothese bestätigen.⁹

In dem vorliegenden Beitrag will und kann ich als Computerlinguist (mit einer Bereitschaft, das computerlinguistische Modell- und Methodeninventar jederzeit zu erweitern oder zu ergänzen, um Fragestellungen aus einem grundsätzlich anderen disziplinären Kontext gerecht zu werden) nicht eine weitere umfassende Beleuchtung des Spannungsfeldes beisteuern. Ich will jedoch im ersten Teil (Abschn. 1.1–1.4) auf einen Aspekt der empirisch-korpusbasierten Methodik abheben, der in der Tat Zweifel daran zu begründen scheint, dass sich diese Methodik in ihrer etablierten Form in den Kern der hermeneutisch-historisch orientierten Literaturwissenschaft – die Textinterpretation – integrieren lässt. Es handelt sich um Fragen der (statistischen) Repräsentativität der Daten, auf deren Grundlage Modellparameter für zukünftige Vorhersagen abgeschätzt werden: Sind diese sogenannten ‚Trainingsdaten‘ – hier also die Texte, deren relevante Eigenschaften exemplarisch von Hand annotiert werden – repräsentativ in Bezug auf die ‚Anwendungstexte‘, d. h. diejenigen Texte, die im Zentrum der eigentlichen Untersuchung stehen? Gewisse inhärente Repräsentativitätsannahmen, die beim Einsatz von Analysemodellen für linguistische Fragen oder Inhaltsrecherchen auf Gebrauchstexten unproblematisch sind, stoßen bei interpretationsrelevanten Eigenschaften literarischer Texte an ihre Grenzen.

Im zweiten Teil (Abschn. 2.1–2.4) argumentiere ich, dass es zwei mögliche Reaktionen gibt, mit diesem methodischen Einwand umzugehen: Die naheliegende, die ich eine ‚defensive Replik‘ nenne, besteht darin, für den Anwendungsbereich der statistischen Methoden nur solche Texte und Texteeigenschaften zu wählen, für die die Repräsentativitätsfrage weitgehend unstrittig ist. Damit wird in Kauf genommen, dass möglicherweise nur ein mittelbarer Beitrag zu Fragen der literarischen Textinterpretation geleistet werden kann (andererseits erschließen sich mit einem korpusorientierten Vorgehen neuartige Fragen, die für eine systematische, historisch orientierte Literaturwissenschaft von großem Interesse sind). Zur Sinnhaftigkeit dieses Vorgehens, das ich in Abschn. 2.1 skizziere, scheint in der Community der *Digitalen Literaturwissenschaft* weitgehender Konsens zu bestehen.

⁹Vgl. Weitin (Anm. 7), 655. Die Gefahr eines umgekehrten Bias, der bei einer aggregierenden Darstellung von Analyseergebnissen für sehr viele Texte den Blick übermäßig stark auf Ausreißer lenkt, spricht Willand (Anm. 7), 85, an.

Alternativ könnte auf den methodischen Einwand jedoch, wie ich in Abschn. 2.2 argumentiere, ‚offensiv‘ reagiert werden, indem der empirische Ausschnitt, für den ein Vorhersagemodell abgeschätzt wird, enger gefasst wird. Nicht intersubjektiv unstrittige Analysekategorien werden modelliert, sondern solche, für die sich unter bestimmten Annahmen eine systematische Beispielannotation spezifizieren lässt – möglicherweise teils subjektiv (etwa wenn ein Typus von impliziten Textaussagen nach einer plausiblen Lesart entsprechend dem subjektiven Leseverständnis systematisch ergänzt werden). Indem man den Fokus auf indirekte empirische Implikationen legt, die sich aus begründeten theoretischen Konzepten ableiten, können korpusbasierte Experimente mit Computermodellen eine empirisch gestützte Auseinandersetzung mit Hypothesen erlauben, die abstraktere Eigenschaften berühren.

Um diesen Gedanken zu konkretisieren, skizziere ich in Abschn. 2.3 einen korpusbasierten Experimentaufbau, der interpretationsabhängige Texteigenschaften als Zielkategorie für statistische Klassifikationsmodelle ansetzt – nämlich einen Aspekt der subjektiven Perspektivierung von Figuren in Erzähltexten. Den Experimentaufbau habe ich für eine Reihe von Texten aus dem Werk Arthur Schnitzlers in einer Pilotstudie umgesetzt, sodass zur Verfügung stehende methodische Möglichkeiten ‚durchgespielt‘ werden können. Gleichwohl muss der Nachweis einer tatsächlichen – hoffentlich gewinnbringenden – Integration in hermeneutische Fragestellungen sicherlich noch erbracht werden.

1 Empirisch-korpusbasierte Methoden und die Interpretation von Texten

1.1 Hintergrund: Empirisch-korpusbasierte Methoden

Text-basierte Explorations- und Analysewerkzeuge lassen sich in vielfältiger Weise für die Aufbereitung und Erschließung digital verfügbarer Texte und ganzer Korpora einsetzen. Einem Einsatz als heuristische Komponente innerhalb einer hermeneutisch-literaturwissenschaftlichen Arbeitspraxis stehen grundsätzlich wenige Beschränkungen entgegen – etwa einem *Topic*-Modell als tentativer Themenfilter beim *Distant Reading* über ein Korpus, mit dem der Blick auf Einzeltexte bzw. Textpassagen gelenkt wird, die anschließend einem *Close Reading* unterzogen werden. Die nachgeordnete Feinanalyse stellt ein methodenkritisches Vorgehen sicher; eventuelle Fehlannahmen bei der Werkzeuganwendung treten hier zutage.

In diesem Beitrag gehe ich jedoch davon aus, dass es innerhalb der Digitalen Literaturwissenschaft ein methodenorientiertes Teilgebiet gibt – vielleicht die *Computational Literary Studies* –, dessen Anspruch und Selbstverständnis über diesen rein heuristisch-explorativen Werkzeugeinsatz hinausgehen: Der Einsatz eines Computermodells soll Ergebnisse erzeugen, die sich direkt in eine Argumentation einfügen lassen – naheliegend sind quantitative Befunde zu Texteigenschaften, die eine studienrelevante These untermauern. Nicht zuletzt um

der Gefahr einer Pseudo-Objektivität zu entgehen, muss die Validität der eingesetzten Modelle stets kritisch geprüft werden – besteht doch aufgrund des zumeist komplexen Zusammenspiels verschiedener Analyseebenen ein Risiko von Scheineffekten (so können etwa Unterschiede in der quantitativen Verteilung eines Schlüsselbegriffs zwischen zwei Korpora ihre Ursache in trivialen orthographischen Besonderheiten haben, mögen aber zu einer inhaltlich begründeten Fehlinterpretation einladen).

1.1.1 Statistische Modellierung in der Korpuslinguistik und Computerlinguistik

Die werkzeuggestützte Korpuslinguistik und die Computerlinguistik¹⁰ arbeiten üblicherweise mit einer Evaluierungsmethodik, die es erlaubt, die Fehlerwahrscheinlichkeit einer Werkzeugkette abzuschätzen, um so die Verlässlichkeit eines Analyseschrittes zu belegen, der beispielsweise Textinstanzen nach einer theoretisch fundierten und empirisch operationalisierten Taxonomie kategorisiert. Wenn etwa anhand eines manuell annotierten Testkorpus nachgewiesen werden konnte, dass ein automatischer *Part-of-Speech Tagger* Wortarten mit einer Genauigkeit von über 90 % erkennt, ist es legitim, Schlüsse aus automatisch ermittelten Häufigkeitsunterschieden im Gebrauch von Adjektiven zu ziehen, die jenseits der Fehlerwahrscheinlichkeit liegen.

Allerdings basiert die Methodik auf Annahmen zu statistischen Eigenschaften der Korpus-daten – insbesondere zum Verhältnis zwischen den Evaluationsdaten und den ‚Anwendungsdaten‘ –, die möglicherweise nicht in jedem Szenario erfüllt sind. Üblich ist der Aufbau eines Basiskorpus, das die Gesamtheit der möglichen Beobachtungen approximiert, aus dem zufällig eine Auswahl von Instanzen gezogen wird, für die dann mehrere geschulte Annotierende unabhängig voneinander die kontextangemessene Zielkategorie von Hand zuweisen. Hierfür muss eine Operationalisierung der Analyseebenen vorliegen, die intersubjektiv nachvollziehbare Entscheidungen ermöglicht. Nach einem Abgleich der Ergebnisse aus der Mehrfachannotation (bei dem konfligierende Analysen entsprechend einem definierten Protokoll aufgelöst werden) kann eine sogenannte *Gold-Standard*-Annotation erzeugt werden, die im weiteren Verlauf näherungsweise als empirische Referenz (*Ground Truth*) betrachtet wird.

Auf einem Teil des *Gold-Standard*-annotierten Referenzkorpus kann schließlich mit überwachten Lernverfahren ein automatisches Klassifikationsmodell ‚trainiert‘ werden, das aus den Eigenschaften der beobachteten Daten und der statistischen Verteilung eine Funktion induziert, die auch für neue Daten eine Vorhersage der jeweils wahrscheinlichsten Zielkategorie erlaubt. Eine zentrale Annahme besteht darin, dass die Daten im Anwendungsfall derselben

¹⁰Eine eingehende Diskussion der Korpus-Annotationsmethodologie in einem sprachtechnologischen Kontext findet sich bspw. in Eduard Hovy/Julia Lavid, „Towards a ‚science‘ of corpus annotation. A new methodological challenge for corpus linguistics“, in: *International Journal of Translation* 22/1 2010, 13–36.

Grundgesamtheit entspringen wie die ‚Trainingsdaten‘. Bei komplexen Aufgaben wird der Klassifikator nie perfekt sein. Das Fehlerrisiko kann jedoch statistisch eingegrenzt werden, wenn ein Teil der *Gold-Standard*-Daten zu Testzwecken aus dem Training herausgelassen wird.

Da gerade in der (Computer-)Linguistik die Erstellung von *Gold-Standard*-Annotationen für anspruchsvolle Analyseaufgaben sehr aufwendig ist (und die meisten Induktionsverfahren eine vergleichsweise große Menge von Trainingsdaten benötigen, um robuste Vorhersage zu liefern), widmet sich ein wichtiger Zweig der angewandten Forschung zu maschinellen Verfahren sogenannten Domänenadaptionstechniken:¹¹ Liegen für das gewünschte Untersuchungskorpus keine – oder nur unzureichend umfangreiche – *Gold-Standard*-Annotationen vor, werden hierbei andere Korpora für das Training eines Ausgangsmodells herangezogen, die hinreichend ähnlich sind. Die resultierenden Modellparameter werden anschließend mit speziellen Verfahren für die Anwendung auf das eigentliche Zielkorpus angepasst. Für eine Validitätsprüfung des resultierenden adaptierten Klassifikators muss bei der Domänenadaption ein Ausschnitt des Zielkorpus als Testdatensatz manuell annotiert werden.¹²

1.1.2 Sind die methodischen Annahmen im literaturwissenschaftlichen Kontext erfüllt?

Generell kann die geschilderte methodische Praxis auf den Einsatz von Computermodellen in literaturwissenschaftlichen Studien übertragen werden – nicht selten kommen existierende sprachtechnologische Modelle oder Modellkomponenten als Basis für ein literaturwissenschaftlich motiviertes Analysemodell infrage (so kann etwa ein generisches Modell für die Erkennung von Eigennamen in Texten die Basis für die Figurenerkennung in Erzähltexten darstellen, evtl. in Kombination mit einem Koreferenzauflösungsverfahren¹³). Wie aus den jeweils angerissenen Annahmen zum Status des Basiskorpus, der manuellen Annotation usw. hervorgeht,

¹¹ Der Terminus ‚Domäne‘ wird in diesem Zusammenhang gerne in einem weiten Sinn in Bezug auf alle Eigenschaften interpretiert, in denen das Korpus der Anwendungstexte von dem der Entwicklungstexte abweicht, also neben der inhaltlichen Domäne, in denen sich die Texte bewegen, auch Textgenre, Stilebene und sogar Sprachstadium u. a. Eigenschaften (vgl. auch Jonas Kuhn/André Blessing, „Die Exploration biographischer Textsammlungen mit computerlinguistischen Werkzeugen – methodische Überlegungen zur Übertragung komplexer Analyseketten in den Digital Humanities“, in: Ágoston Z. Bernád/Christine Gruber/Maximilian Kaiser (Hg.), Europa baut auf Biographien. Aspekte, Bausteine, Normen und Standards für eine europäische Biographik, Wien 2018, 225–257).

¹² Das skizzierte Vorgehen betrifft überwachte Lernverfahren; daneben können Analysemodelle regelbasiert erstellt werden, oder es können unüberwachte Lernverfahren eingesetzt werden, die Muster in der Datengrundlage automatisch erfassen. Auch in diesen Fällen muss für eine methodisch fundierte Modellanwendung eine Validierung anhand von operationalisierten Zielkategorien durchgeführt werden – ein repräsentatives Evaluationskorpus mit *Gold-Standard*-Annotation muss also unabhängig vom Modellierungsverfahren vorliegen bzw. zur Absicherung der Validität erzeugt werden.

¹³ Einschränkungen werden in Jannidis u. a. (Anm. 2) diskutiert.

liegt dem Arbeitsablauf jedoch ein bestimmtes wissenschaftstheoretisches Bezugssystem zugrunde: die Analyse oder Annotation eines Texts mit dem Ziel, seinen Teilinstanzen bestimmte Kategorien zuzuweisen, wird im Grunde als ein Experiment betrachtet, für das unter gleichen Kontextbedingungen reproduzierbare Ergebnisse angenommen werden. Diesem naturwissenschaftlich geprägten Bezugssystem mag sich nicht unbedingt jede Hermeneutikerin oder jeder Hermeneutiker anschließen. Allerdings dürfte einleuchtend sein, dass für einen fundierten Einsatz einer algorithmischen Komponente mit bestimmten Ein- und Ausgabestrukturen deren Repräsentativität geklärt sein muss.

Wenn man sich den eben angesprochenen grundsätzlichen Bedenken nicht anschließt (und davon wird nachfolgend ausgegangen), muss also weiterhin geprüft werden, inwiefern die methodischen Arbeitshypothesen bei einem literarischen Gegenstand und literaturwissenschaftlichen Analysezielen tragfähig sind. Hierbei geht es nicht darum, ob sämtliche methodisch motivierten Annahmen zum Untersuchungsszenario faktisch immer erfüllt sind – im Zuge einer theoriegeleiteten wissenschaftlichen Agenda wird es in vielen Fächern als legitim angesehen, idealisierende Annahmen zu machen, sofern argumentiert werden kann, dass der Zugang zur Empirie dadurch *grosso modo* unverfälscht bleibt –, wenn etwa für bestimmte Experimente in der Mechanik bewusst die Reibung ignoriert wird. Es reicht dann also, wenn geringfügige Abweichungen von den Annahmen ohne Einfluss auf wesentliche Aspekte der Fragestellung sind – wobei Einigkeit in der Community darüber bestehen muss, welcher Abstraktionsschritt die Idealisierung trägt.¹⁴

Will man nun den korpusbasierten statistischen Analyseansatz innerhalb der hermeneutisch-literaturwissenschaftlichen Praxis anwenden, gibt es nach meinem – in vieler Hinsicht sicherlich unvollständigen – Verständnis mehrere Punkte, in denen man sich davon überzeugen muss, dass die geschilderten methodischen Annahmen in ausreichendem Maße erfüllt sind. Auf zwei Punkte will ich ausführlicher eingehen: zum einen auf die Repräsentativitätsannahme, u. a. in der Beziehung zwischen dem Basiskorpus und einzelnen Untersuchungsgegenständen (dazu Abschn. 1.3), zum anderen die Möglichkeit der Festlegung auf eine Referenzlösung (*Ground Truth*) für zentrale Analysekatoren (Abschn. 1.4).

¹⁴Dies lässt sich freilich nicht unabhängig von spezifischen theoretischen Vorannahmen beurteilen. Die Arbeitspraxis der modernen Linguistik kennt bspw. sehr weitreichende Idealisierungen – vor allem zu theoretischen Konstrukten, die nicht direkt empirisch beobachtet werden können (und hierzu gehören bereits zentrale grammatische Beschreibungskategorien wie der Begriff des Subjekts oder die Struktureinheit eines Wortes). Zu deren arbeitspraktischer Legitimation gehört es in der Regel, dass mehrere unabhängige operationalisierte empirische Tests zu einer konsistenten Kategorisierung führen müssen. Zwar ist und war die Frage, wie weit theoriespezifische Annahmen im Rahmen dieser Praxis gehen dürfen, Gegenstand vehementer disziplininterner Auseinandersetzungen (bspw. die Kritik an Aspekten der theoretischen Ansätze von Noam Chomsky – etwa der Annahme von nicht direkt beobachtbaren Transformationsoperationen, vgl. u. a. Joan Bresnan, *Lexical-Functional Syntax*, Oxford 2001). Auf höheren Abstraktionsebenen besteht jedoch weitestgehend Konsens, dass idealisierte Konstrukte einem systematischen Zugang zur Empirie zuträglich sind.

1.2 Repräsentativität der Korpusauswahl

Bereits das Auswahlprinzip für die Zusammenstellung des Basiskorpus, das als Approximation der theoretisch anzunehmenden Grundgesamtheit von *möglichen Texten*¹⁵ fungiert, ist im literarischen Kontext nicht stärker virulent als für die Linguistik und Sprachtechnologie, die sich seit Jahrzehnten mit pragmatischen Lösungen behelfen: Für sehr wenige Sprachen liegen ausreichend große balancierte Korpora vor (bei denen der Versuch gemacht wird, die relative Verteilung von Quellen für Sprachinput, denen Muttersprachler ausgesetzt sind, zu spiegeln). Stattdessen werden notgedrungen zumeist die Korpora genutzt, die in der jeweiligen Untersuchungssprache eben verfügbar sind, nicht selten (möglichst große) Zeitungskorpora. Empirische Aussagen können dann nicht über den gesamten Sprachgebrauch gemacht werden, sondern nur relativ zum realen Basiskorpus. Eine solche Vorauswahl ist für die (Computer-)Linguistik unter der idealisierenden Annahme tragbar, dass die Verteilung von strukturellen grammatischen Phänomenen weitgehend unabhängig von Gattungsspezifika bzw. der Inhaltsdomäne etc. ist. Dies ist oft nicht der Fall – z. B. sind direkte Fragesätze in Zeitungstexten stark unterrepräsentiert (was in der Praxis dazu

¹⁵ Es mag zunächst fraglich erscheinen, ob für literaturwissenschaftliche Untersuchungen überhaupt ein Ansatz gewählt werden sollte, der mit inferenzieller Statistik arbeitet und die vorhandenen Texte als Stichprobe aus einer größeren Grundgesamtheit betrachtet. Dieses Vorgehen ist für die Linguistik angemessen, welche das Sprachsystem modelliert, das nicht nur die tatsächlich realisierten Äußerungen umfasst: Kompetente Muttersprachler sind in der Lage, die Grammatikalität bzw. die möglichen Bedeutungen von denkbaren, jedoch nie realisierten Äußerungen zu beurteilen. Die Aufgabe der Literaturwissenschaft beschränkt sich jedoch auf die Betrachtung von realen Texten. Daher könnte man annehmen, dass es ausreicht, mit Verfahren der deskriptiven Statistik zu arbeiten – also keine Grundgesamtheit hypothetisch möglicher Texte anzusetzen. Aus verschiedenen Gründen ist es jedoch in Teilbereichen der Digitalen Literaturwissenschaft sinnvoll, den konzeptionellen Rahmen der inferenziellen Statistik zu verwenden. Zu direkt messbaren Eigenschaften der untersuchten Texte (etwa der Länge in Druckseiten, Kapiteln oder Zeichen) oder manuellen Annotationen können selbstverständlich deskriptive Statistiken berechnet werden. Wenn jedoch ein Werkzeug eingesetzt wird (etwa bereits ein einfaches Regelsystem zur Bestimmung von Satzgrenzen oder anspruchsvollere linguistische oder analysespezifische Werkzeuge), kann eine Qualitätsüberprüfung in der Regel nur anhand von Stichproben vorgenommen werden, und die Frage wird virulent, wie stark die Texte im Anwendungsfall variieren. Darüber hinaus kann es auch trotz eines fixen Korpus (eventuell sogar der vollständigen Überlieferung aus einer Epoche) sinnvoll sein, Entwicklungen mit Hilfe von Modellen der inferenziellen Statistik zu modellieren. Wenn etwa die Hypothese geprüft werden soll, dass sich ein Autor mit einem Text zu einem bekannten Stoff vom Epochenstil absetzt, kann mit einem inferenziell-statistischen Ansatz viel feingliedriger zwischen einem hypothetischen stilkonformen Text zu diesem Stoff und dem tatsächlich entstandenen Text verglichen werden. Nicht zuletzt spricht die Möglichkeit, dass verschollene Texte wiederentdeckt werden, Nachahmungen zu überprüfen sind oder im Rahmen der empirischen Rezeptionsforschung Varianten von Texten synthetisiert werden können, dafür, dass die Annahme eines prinzipiell unendlichen Raums möglicher Texte auch in Teilen der Literaturwissenschaft sinnvoll sein kann.

führt, dass Syntax-Parser, die auf Standardkorpora trainiert wurden, bei Fragen tendenziell mehr Fehler machen). Für viele Untersuchungen zu grammatik-internen Interaktionen – etwa zwischen unterschiedlichen Auslösern für Wortstellungspräferenzen (Subjekt-Präzedenz vs. Präzedenz von Pronomina gegenüber vollen Phrasen) – kann jedoch ein beliebiger Sprachausschnitt eine durchaus verwertbare empirische Grundlage liefern: Da sich die (Computer-)Linguistik mit Regularitäten in der Sprache als ein System befasst, das (einer linguistischen Grundannahme nach) jeder Textproduktion und -perzeption zugrunde liegt, sind die Hürden der Basiskorpuskonstruktion in der Praxis nicht allzu hoch, und verfügbare Korpora lassen sich zumeist für sehr viele unterschiedliche Studien nutzen. Aus diesem Zusammenhang folgt auch, dass Modelle, die aus einem Ausschnitt des Basiskorpus entwickelt wurden, bei der Anwendung auf völlig neue Sätze der fraglichen Sprache in aller Regel zu halbwegs robusten Analyseergebnissen führen – dank der grammatisch-strukturellen Verwandtschaft.

Auf literaturwissenschaftliche Untersuchungen übertragen, könnte für oberflächennahe Texteigenschaften in Korpora, die eine sehr große Zahl von ganzen Texten enthalten, langfristig möglicherweise eine ähnliche Praxis etabliert werden.¹⁶ Für typische komplexere Analysefragen jedoch, die für einen konkreten literarischen Text oder ein kleines Korpus zu bearbeiten sind, scheint die Bezugsdimension zu einem relevanten Vergleichskorpus von Studie zu Studie stark zu variieren: Soll betrachtet werden, inwiefern der Text zeitgenössische Gattungskonventionen (bzw. den Kanon) widerspiegelt und wo er sie durchbricht? Versucht eine produktionsästhetische Studie zur empirischen Untermauerung eines Interpretationsansatzes Tendenzen in Briefen und Tagebucheinträgen der Autorin oder des Autors zu finden? Können intertextuelle Bezüge zu Texten anderer Autorinnen und Autoren, teilweise aus ganz anderen Epochen, sichtbar gemacht werden? In jedem Fall ist das Auswahlprinzip zur Zusammenstellung eines geeigneten Basiskorpus ein grundsätzlich anderes. Diese Vielfalt erschwert nicht nur ganz praktisch die Übertragbarkeit von (in aller Regel aufwändig erstellten) Korpusressourcen – die jeweils unterschiedlich gelagerten Prinzipien dürften auch häufig den Einsatz bekannter Modellierungsverfahren verhindern. Prinzipiell wäre es denkbar (und forschungsökonomisch effizient), dass über mehrere Arbeitsgruppen zur *Digitalen Literaturwissenschaft* hinweg eine Abstimmung der Agenda zu Korpusarbeiten und der Modellierung von Analyseaufgaben stattfindet, sodass Synergie-Effekte

¹⁶Ein funktionierendes Beispiel für eine ergiebige Abstraktion über sehr heterogene Textsammlungen hinweg sind stilometrische Verfahren wie John Burrows's „Delta“. A Measure of Stylistic Difference and a Guide to Likely Authorship“, in: *Literary and Linguistic Computing* 17 (2002), 267–287. Mit ihnen wird wie in der Linguistik eine systematische Ebene erreicht (in diesem Fall die relative Verteilung der häufigsten Wortformen im Stil eines Autors – also eine Beobachtung auf der Meta-Textebene), die robust messbar und auf neue Texte (oder Text/Autor-Paare) übertragbar ist.

ausgenutzt werden können.¹⁷ Dies läuft allerdings der etablierten Publikationspraxis und dem Originalitätsgedanken entgegen und könnte daher eher ein Ziel für einen langfristigen Prozess des Wandels sein.

Setzt man vor dem Hintergrund der zuletzt diskutierten Überlegungen Aufwand und Ertrag für den Einsatz von anspruchsvollen korpusbasierten Methoden in Relation – für Analysen, die einen Kernbeitrag zu einer relevanten Leitfrage leisten können –, könnten auf absehbare Zeit konventionelle Praktiken häufig noch den Vorzug erhalten.

In jedem Fall scheint sich die Konstellation aus Gegenständen und typischen Fragestellungen in den Literaturwissenschaften so darzustellen, dass es (noch?) unrealistisch ist, ein Korpus für die studienübergreifende Unterstützung von ‚tieferen‘ Analysefragen zusammenzustellen. Der repräsentative Charakter, den bei *linguistischen* Untersuchungen ein klug gewähltes sprachspezifisches Textkorpus für sehr viele Analysefragen bietet, lässt sich durch ein generisches Korpus mit literaturwissenschaftlicher Breite sicherlich nur in geringerem Maß erreichen. Hierzu müsste der jeweils untersuchte Textgegenstand unter der idealisierenden Annahme betrachtet werden, dass er zu der angenommenen Grundgesamtheit gehört, die durch das Korpus von Anfang an approximiert wurde. Dies läuft jedoch einer verbreiteten literaturwissenschaftlichen Grundmotivation entgegen, nach der durch originelle Schritte der Textanalyse das Singuläre herauszuarbeiten ist, das einen Text auszeichnet, und textspezifischen ästhetischen Aspekten gerecht zu werden.¹⁸ Für eine Vielzahl von Studien muss also auf absehbare Zeit in einen fragespezifisch motivierten Korpusaufbau investiert werden.

1.3 Die Möglichkeit einer *Gold-Standard-Annotation* für zentrale Analysekategorien

Die geschilderte Problematik der Repräsentativität von vorhandenen Korpora für diese oder jene Kombination aus Gegenstand und Fragestellung ist im Grunde auf eine Abwägung von Aufwand und Ertrag zurückzuführen: Auch für einen sehr spezifischen Blickwinkel *könnte* prinzipiell ein repräsentatives Korpus konstruiert werden. Aus nachvollziehbaren Gründen wird sich dies jedoch nicht (rasch) zur Standardpraxis in der Breite des Fachs entwickeln – liegen doch andere anerkannte Argumentationsmuster vor, mit denen Thesen zum relevanten

¹⁷ Überlegungen zu einem entsprechenden Umgang mit der Problematik der Zeitplanung bei Digital Humanities-Projekten (dem Scheduling Problem) diskutiere ich in Jonas Kuhn, „Computational text analysis within the Humanities. How to combine working practices from the contributing fields?“, *Language Resources and Evaluation*, im Erscheinen.

¹⁸ Was hingegen praktikabel erscheint, ist eine differenzielle Bestimmung der Dimensionen, in denen ein gegebener Text ‚Ausreißer‘ gegenüber Mustern darstellt, die im Korpus manifest sind. Vgl. auch die Diskussion von Ausreißern bei der Interpretation von Visualisierungen bei Willand (Anm. 7), 85. Ein entsprechendes ComputermodeLL schiene mir jedoch eher den Charakter eines peripheren Explorationswerkzeugs zu haben.

Kontextbezug eines Texts begründet werden können. Nun kommen wir zu einer grundlegenden Problematik: Ist die Notwendigkeit einer Festlegung von Referenzlösungen im *Gold Standard* (evtl. für Trainingsdaten, in jedem Fall aber für die Modellvalidierung) mit dem Selbstverständnis eines hermeneutischen Ansatzes in der Literaturwissenschaft vereinbar? Auf welcher Basis soll die Entscheidung für eine intersubjektiv akzeptierte, ‚korrekte‘ Annotation gefällt werden, wenn unterschiedliche Textinterpretationen zu konkurrierenden Ergebnissen führen? Ich will hier nicht die Debatte zum Status der Polyvalenz-These (dass die Vieldeutigkeit ein zentrales – vielleicht konstitutives – Merkmal literarischer Texte sei) nachvollziehen¹⁹ – dazu fehlt mir auch die literaturwissenschaftliche Kompetenz. Es lohnt sich jedoch, die Dimensionen der Interpretationsproblematik abzustecken, um zu verstehen, für welche Art von analytischer Fragestellung welcher Ansatz systematisch-strukturell geeignet ist.²⁰ Insofern sich die *Digitale Literaturwissenschaft* geeigneter Analysemodelle aus der Computerlinguistik bedienen kann, bietet sich wiederum ein knapper Vergleich mit der Annotationspraxis für die linguistische Textanalyse an.

Nicht wenige Analyseentscheidungen, die bei einer linguistischen Korpusannotation gefällt werden müssen, erfordern eine (linguistische) Interpretation des Satzes im gegebenen Kontext, sprich eine Rekonstruktion der Inferenzschritte, die zur Zuweisung der nächstliegenden kontextangemessenen Bedeutung des Satzes führen. Dies trifft – zunächst vielleicht überraschend – nicht nur für lexikalisch-semantische und satz-semantische Entscheidungen zu (etwa: welche Lesart von ‚einstellen‘ ist gemeint? „Maschinen einstellen“/„Lehrer einstellen“), sondern auch für rein grammatische Annotationen: Um Subjekt- und Objektstatus in einem Satz wie

¹⁹Jannidis argumentiert überzeugend gegen ein Verständnis der Polyvalenz-These, nach dem es unmöglich sei, zwischen zwei konkurrierenden Textinterpretationen argumentativ zu unterscheiden, vgl. Fotis Jannidis, „Polyvalenz – Konvention – Autonomie“, in: Ders./Gerhard Lauer/Matías Martínez u. a. (Hg.), *Regeln der Bedeutung. Zur Theorie der Bedeutung literarischer Texte*, Berlin/New York 2003, 305–328. Die Festlegung einer Gold-Standard-Lösung für interpretationsrelevante Analyseentscheidungen in einem Referenzkorpus ginge jedoch einen Schritt weiter, indem diese ein für alle Mal eine Variante zur Referenz erhebt (zumindest, wenn zur Annotationsentscheidung nicht die Relevanzkriterien abgelegt werden, die genau zu diesem Schluss geführt haben – was im größeren Stil wohl nicht praktikabel wäre).

²⁰An anderer Stelle habe ich vor dem Hintergrund der Diskussion zur Annotationspraxis in Computerlinguistik und (Digitaler) Literaturwissenschaft bereits ähnliche Überlegungen angestellt, vgl. dazu Adam Hammond/Julian Brooke/Graeme Hirst, „A tale of two cultures: bringing literary analysis and computational linguistics together“, in: *Proceedings of the NAACL 13 Workshop on Computational Linguistics for Literature*, Atlanta, GA, 2013, 1–8; Jonas Kuhn, „Computerlinguistische Textanalyse in der Literaturwissenschaft? – oder: ‚The Importance of Being Earnest‘ bei quantitativen Untersuchungen“, in: Andrea Albrecht/Sandra Richter/Marcel Lepper (Hg.), *Quantitative Verfahren in der Literaturwissenschaft. Von einer Scientia Quantitatis zu den Digital Humanities*, Berlin/Boston 2018, 11–44.

1. Diese Einschätzung teilt die Opposition nicht.

zuweisen zu können, genügt es nicht, die Kasusmarkierung zu überprüfen (die Formen sind synkretistisch), und die Wortstellung gibt im Deutschen nicht den Ausschlag. Fehlt Wissen zum Interpretationskontext, kann dies zu Fehlanalysen auf der grammatischen Ebene führen: Nehmen wir an, wir wüssten nicht, dass Mangusten eine Marderart sind, sondern wir wären der Annahme, es handle sich um Insekten. Dann würden wir in folgendem Satz Subjekt und Objekt vermutlich falsch zuordnen:

2. Auch die große Kröte erwischte die Manguste nach geduldigem Lauern.

Linguistische Annotationsaufgaben bewegen sich also immer im Spannungsfeld von unterschiedlichen Interpretationen. Die beiden angeführten Satzbeispiele sind grammatisch gesehen ambig (man spricht von ‚globaler‘ Ambiguität, da jeweils die ‚lokale‘ Nominativ/Akkusativ-Ambiguität der beiden Nominalphrasen auch im globalen Satzzusammenhang nicht grammatisch aufgelöst wird, wie dies sonst zumeist der Fall ist). Es wäre denkbar, solche Sätze bei der *Gold-Standard*-Annotation für die Analysen von grammatischen Relationen als ‚unaufgelöst ambig‘ zu markieren (in der Grammatiktheorie wird hierfür das Konzept der ‚Unterspezifikation‘ verwendet). Dann würde die Annotation nicht durch außergrammatische Informationen ‚kontaminiert‘.

Die übliche Annotationspraxis erfordert jedoch bei strukturell ambigen Instanzen eine Entscheidung – auf Basis des semantischen (und ggf. pragmatischen) Verständnisses. Dies ist sinnvoll, da die überwiegende Zahl von globalen Ambiguitäten von kompetenten Sprechern/Hörern überhaupt nicht wahrgenommen wird – das semantische Mitlesen können wir nicht abschalten (strukturelle Ambiguitäten wie in Beispiel (1) dürften die meisten Leser bei der inhaltlichen Lektüre übersehen²¹). Und da die Interaktion zwischen unterschiedlichen Granularitätsebenen (lexikalische Semantik, syntaktische Struktur, semantische Skopusentscheidungen, Informationsstruktur im Sinn von Fokus/Hintergrundgliederung) sehr komplex ist, wäre eine Spezifikation sämtlicher Umstände, unter denen Ambiguitäten explizit markiert werden sollen, sehr kompliziert und dürfte in der Praxis zu recht arbiträren Einschätzungen führen. Die Aufforderung an die Annotierenden, die im Kontext jeweils plausible Interpretation zugrunde zu legen und danach alle Teilentscheidungen aufzulösen, ist hingegen simpel und robust. Der wahrscheinlich größte Vorteil liegt wohl darin, dass die entstehende *Gold-Standard*-Annotation als Basis für Analyseentscheidungen auf sehr vielen Ebenen herangezogen werden kann. So können morphologische Annotationen innerhalb einer komplexen syntaktischen Struktur extrahiert werden, um einen morphologisch angereicherten *Part-of-Speech Tagger* zu trainieren, der

²¹ Der Satz in dieser Klammer ist ein weiteres Beispiel.

innerhalb von lokal ambigen Phrasen ggf. lexikalisch-semantiche Tendenzen berücksichtigt (der also in einem Satzanfang „Diese Einschätzung teilt ...“ bereits ein Akkusativobjekt vermutet, obgleich der Tagger keine vollständige syntaktische Struktur aufbaut).

Die Praxis der erzwungenen (kontextgebundenen) Disambiguierung stößt in zwei Situationen an Grenzen: zum einen bei Sprachspielen, die die Ambiguität auf die Ebene des Bewussten ziehen, wie folgender Verwechslungswitz auf Basis der lexikalischen Ambiguität von „Schale“:

3. Eine Frau kommt zum Arzt und sagt: „Mein Kind hat eine Apfelsine mit der Schale gegessen, ist das schlimm?“
 Darauf der Arzt: „Wenn die Schale nicht gespritzt war, ist das halb so wild.“
 Die Frau: „Gespritzt? Die Schale war aus Porzellan!“²²

In üblichen Korpora, die für linguistische und sprachtechnologische Zwecke eingesetzt werden, kommen solche Sprachspiele in verschwindend geringer Zahl vor. (Wobei die erzwungene Disambiguierung wohl korrekterweise den Zeitpunkt der Pointe vorhersagen würde: Es würde im ersten Satz die naheliegende Interpretation – des Arztes – annotiert, die erst im letzten Satz scheitert; die Auflösung der Pointe aufgrund der alternativen Interpretation kann jedoch aus der Annotation nicht rekonstruiert werden).

Die zweite problematische Situation tritt dann ein, wenn unterschiedliche Annotierende bei der kontextgebundenen Interpretation zu unterschiedlichen Ergebnissen kommen, die jedoch orthogonal zur eigentlichen Annotationsaufgabe stehen und daher nicht explizit in der Annotation dokumentiert werden. Dies kann – selten – zu irreführenden Effekten führen. Glauben bei der Annotation von Beispiel (2) etwa drei von vier Annotierenden fälschlich, dass Mangusten kleine Insekten sind, könnten sich Fehler in das *Gold-Standard*-Korpus einschleichen, die das oben angedeutete ‚vorausahnende‘ Verhalten von Modellen beeinträchtigen könnten. Diese Möglichkeit führt jedoch im üblichen Rahmen einer linguistischen Annotation zu keinen systematischen Problemen: Die Aufgabe der grammatischen Kernannotation greift auf die Sprachkompetenz zu, und der interpretationsrelevante Disambiguierungshintergrund beschränkt sich bei den üblicherweise annotierten Texten auf das Allgemeinwissen von Zeitungslesenden o. ä. In Einzelfällen werden Annotationsfehler auftreten; diese sind jedoch nicht systematisch verteilt und führen daher höchstwahrscheinlich zu keiner Verfälschung bei der Modellentwicklung oder -validierung. Mit anderen Worten: Für die linguistische Annotation gewährleistet ein Bezug auf muttersprachliche Sprachkompetenz in Kombination mit Inferenzen aufgrund von Allgemeinwissen eine robuste Basis für intersubjektiv stabile, voll disambiguierte Annotationen.

²²Wikipedia, <http://de.wikipedia.org/wiki/Verwechslungswitz> (letzter Aufruf 22.09.2017).

1.4 Literaturwissenschaftliche Interpretation und intersubjektive Annotation

Kommen wir nun zum literaturwissenschaftlichen Rahmen. Über die linguistischen Analysefragen hinaus werden hier beispielsweise Analysekatgorien aus der narratologischen Beschreibung relevant. Die zu annotierenden Texte sind literarische Texte. Wie verändert sich die systematisch-strukturelle Ausprägung der Annotationsaufgabe? Zu der Ebene der oben (und in der linguistischen Literatur) so genannten linguistischen Interpretation tritt die Ebene der literaturwissenschaftlichen Interpretation hinzu. Folgt man der Terminologie eines Eric D. Hirsch Jr., kommt zur inhaltlichen Bedeutungsebene des Texts (*Meaning*) seine (mögliche) Bedeutung im Sinne von Bedeutsamkeit (*Significance*) hinzu.²³

Die Frage, ob sich der Blick auf die inhaltliche Bedeutungsebene gegenüber der bisherigen, linguistischen Betrachtung im Rahmen der Korpusannotation für Literaturstudien verschieben kann bzw. muss, wird uns weiter unten eingehender beschäftigen (Abschn. 2.3). Fürs Erste ist festzustellen, dass für Analyseentscheidungen, die in engem Zusammenhang mit der Ebene der literaturwissenschaftlichen Interpretation stehen, ohne weitere Qualifizierung keine intersubjektiv stabile Annotation zu erwarten ist. Dies betrifft etwa die Fragen, ob eine Schlüsselszene im Text allegorisch ist, was die psychologische Deutung einer bestimmten Wendung ist, worauf der intertextuelle Bezug zu einem Text anspielt usw. Auch ohne die stärkeren Implikationen der Polyvalenz-These ist unstrittig, dass für denselben Text – abhängig vom als relevant betrachteten Kontext – mehrere divergierende literaturwissenschaftliche Interpretationen nebeneinander stehen können (häufig eingebettet in unterschiedliche literarische Interpretationstheorien). Es erscheint auch nicht realistisch, Annotierende in den Richtlinien (*Guidelines*) eine kompakte Charakterisierung des relevanten Interpretationskontexts an die Hand zu geben, der in der Mehrzahl der Fälle zu einer intersubjektiven Konvergenz führte. (Und selbst wenn dies für bestimmte, stark formalisierte Interpretationskontexte möglich wäre, wären die entstehenden Annotationen für andere Kontexte von sehr eingeschränktem Wert.) Ein Einsatz der etablierten Korpusannotationspraxis, die auf überindividuelle Konvergenz abzielt, kommt also auf oberster Ebene der Literaturinterpretation nicht infrage.

Wie verhält es sich aber mit weniger komplexen Entscheidungen bei der Analyse von literarischen Texten? Kann hier durch Parallelannotation ein intersubjektiver Konsens ermittelt werden und dann für (das Training und) die Validierung von Vorhersagemodellen herangezogen werden? Hierzu müssen wir die Interaktion zwischen Analyseentscheidungen auf verschiedenen Abstraktionsebenen betrachten. Gibt es den Effekt eines ‚Ausstrahlens‘ von tieferen Entscheidungen auf solche, die näher an der Oberfläche sind – so wie wir es in der

²³ Vgl. Eric D. Hirsch Jr., *Validity in Interpretation*, Yale 1967. Unter *Significance* fasst Hirsch sehr breit die Beziehung zwischen der inhaltlichen Bedeutung des Textes und einer Person oder einer Konzeption, einer Situation „or indeed anything imaginable“ (ebd., 8).

Diskussion der Beispiele (1) und (2) bei der Abhängigkeit von strukturellen Disambiguierungsentscheidungen zur Satzstruktur von der inhaltlichen (linguistischen) Bedeutung beobachtet haben? Selbstverständlich lässt sich dieser Effekt auch beim Zusammenspiel von Literaturinterpretation und deskriptiven Kategorien der Textanalyse nicht abstreiten. Man wird einen Text tendenziell so lesen, dass Passagen, die einen Interpretationsspielraum (im Kleinen) offen lassen, konsistent mit einer sehr weitgehenden, komplexen Interpretation ‚ausgefüllt‘ werden.

Ein dankbares Beispiel für die Manifestation von weitreichenden interpretatorischen Thesen in minimalen sprachlichen Differenzierungen ist sicherlich die Rezeption von Christoph Ransmayrs Roman *Die letzte Welt*,²⁴ der schildert, wie Ovids Freund Cotta diesen – Publius Ovidius Naso – in seinem Exil am Schwarzen Meer aufzufinden versucht, aber nur vielfältige Spuren Ovids findet. Im Verlauf des Romans scheint Cotta sich mehr und mehr mit Naso zu identifizieren. In einer Rezension in *DIE ZEIT* beobachtet Volker Hage dazu:

4. Zweimal läßt der Autor Cotta auf den Spuren Nasos ins Gebirge gehen. „Hier war Naso gegangen: Das war Nasos Weg.“ So heißt es beim ersten Mal. Beim zweiten Mal, am Ende des Romans, eine minimale Verschiebung: „Hier war Naso gegangen; dies war Nasos Weg.“ Ganz unauffällig, fast zu überlesen: eine Ineinssetzung – die Figuren fallen zusammen.

Ist Cotta Naso geworden? War sein Weg, den wir verfolgt haben, auch der des Dichters? Cotta, oder wer es ist, sucht einen letzten Stoffetzen: den mit seinem eigenen Namen. Er selbst ist Bestandteil jener Geschichte geworden, die es nur noch in Fragmenten gibt, die sich in einer leeren, letzten Welt verlieren. Wie mögen die zwei Silben, die auf dieser Fahne stehen, lauten? Cotta? Naso? Oder vielleicht Christoph?

Dieses Rätsel läßt Christoph Ransmayr offen.²⁵

Die Referenz der Demonstrativpronomina *das* und *dies* (welche zunächst auf der rein inhaltlichen Bedeutungsebene [= Hirschs *Meaning*] verortet ist) ist also mit weitreichenden Fragen der Interpretation verbunden.²⁶

Durch die unbestritten vorhandenen, Ebenen übergreifenden Abhängigkeiten ‚sichert‘ der Einwand gegen die Möglichkeit einer intersubjektiv tragbaren Annotation von der Ebene der abstrakten Literaturinterpretation hinunter

²⁴Christoph Ransmayr, *Die letzte Welt*, Roman, mit einem Ovidischen Repertoire, Nördlingen 1988.

²⁵Volker Hage, „Mein Name sei Ovid. Ein großer Roman. Christoph Ransmayrs ‚Die letzte Welt‘“, in: *DIE ZEIT* 41 (1988), <http://www.zeit.de/1988/41/mein-name-sei-ovid/> (letzter Aufruf 22.09.2017).

²⁶Wollte man entsprechende Unterscheidungen als relevante Texteingenschaften annotieren, müsste das Beschreibungsinventar, wie zur Charakterisierung der Verwechslung im Witz (3), natürlich in der Lage sein, eine vorhandene Mehrdeutigkeit zu charakterisieren.

auf einzelne textanalytische Entscheidungen, soweit sie interpretationsrelevant sind. Es kann nicht ausgeschlossen werden, dass zwei Annotierende, die unterschiedliche Gesamtinterpretationen im Kopf haben, allein aus diesem Grund gegenläufige Annotationsentscheidungen vornehmen, und dies konterkariert die Brauchbarkeit der Annotation als *Gold-Standard*.

Augenscheinlich resultiert also das Dilemma, dass nur für solche Analysekatoren eine valide komputationelle Modellierungspraxis etabliert werden kann, deren Wertebereich in keinem direkten Bezug zu Fragen der literaturwissenschaftlichen Interpretation steht. Dies wäre freilich eine Rechtfertigung für ein eher peripheres Interesse an den neuen Methoden innerhalb des etablierten Faches.

2 Reaktionen auf die methodischen Einwände

2.1 ‚Defensive‘ Reaktionen

Methodischen Einwänden, laut denen die übliche Praxis der korpusbasierten Entwicklung und Validierung von Computermodellen bei typischen literaturwissenschaftlichen Konstellationen von Untersuchungsgegenstand und Fragestellung an Grenzen stoße, lässt sich aus Sicht der *Digitalen Literaturwissenschaft* in mindestens zweierlei Weise begegnen: Der erste Typ von Replik, den ich hier (wertungsfrei) ‚defensiv‘ nenne, anerkennt die Problematik eines Ausstrahlens zugrunde liegender literaturwissenschaftlicher Interpretationsansätze auf eine Reihe von Textanalyseaufgaben (für die damit die datengeleitete Standardmethodik der Modellierung nicht infrage kommt), verweist jedoch darauf, dass es eine Vielzahl von *deskriptiven* Analysekatoren gibt, die sich unabhängig von der Textinterpretation operationalisieren lassen – beispielsweise ‚stabile‘ oberflächennahe narratologische Analysekatoren²⁷ oder Texteigenschaften, die sich aus Metadaten zum Text ableiten (wie Autorenschaft, Erscheinungsdatum etc.). Eine systematisierte Modellierungspraxis ermöglicht die Validierung von Analyseinstrumenten und eine Skalierung des Volumens von Text(en), die auf derartige deskriptive Texteigenschaften untersucht werden können. Je breiter die zukünftige Basis derartiger Operationalisierungen und, wo robust möglich, Vorhersagemodelle, desto vielseitiger die Möglichkeiten von differenzierten quantitativen Betrachtungen auf größeren literaturwissenschaftlichen Korpora. Wie bereits die Ergebnisse der letzten Jahre andeuten (mit dem Paradebeispiel stilometrischer Untersuchungen zu Autorschaft und darüber hinaus zu feineren Differenzierungen), dürfte sich das Spektrum der Fragestellungen, die sich fundiert bearbeiten lassen, spürbar erweitern – teilweise sicherlich in Richtungen, die noch nicht absehbar sind. Der Einschränkung auf deskriptive Katoren, die in keinem unmittelbaren Zusammenhang zur Textinterpretation stehen, wird

²⁷Vgl. die Annotations-Guidelines aus dem heureCLÉA-Projekt: Evelyn Gius/Janina Jacke, Zur Annotation narratologischer Katoren der Zeit. Guidelines zur Nutzung des CATMA-Tagsets, Version 2, Hamburg, November 2016.

also wettgemacht (a) durch die Skalierung der Analysemöglichkeit auf ungleich größere Textmengen, die valide und ‚reliabel‘ analysiert werden können, und (b) durch die Möglichkeit der systematischen Kombination von, für sich genommen, wenig komplexen Filtern, die schnell zu nicht-trivialen Beobachtungen auf dem Untersuchungskorpus führen können und so die etablierte Praxis der historisch fundierten Literaturwissenschaft um wertvolle Instrumente erweitern.²⁸ Nicht zuletzt dürfte der Bedarf an operationalisierten Kriterien für Analyseentscheidungen mittelfristig die Auseinandersetzung mit der Begriffsbildung für zentrale Konzepte (wie Gattungsbegriffe, den Status von Kanones etc.) im Kernfach beleben.²⁹

Sicherlich sind auf dem Wege der vielfältigen denkbaren ‚defensiven‘ Reaktionen auf den methodischen Einwand erhebliche methodische Fortschritte zu erwarten (die den Charakter der *Digitalen Literaturwissenschaft* stark prägen dürften). Dennoch werde ich in Abschn. 2.2 argumentieren, dass es eine alternative ‚offensivere‘ Reaktionsmöglichkeit gibt, die parallel dazu verfolgt werden könnte. Als Motivation könnten zwei Überlegungen dazu angeführt werden, weshalb die defensive Reaktion nicht ganz unproblematisch ist.

Zum einen ist die Ausarbeitung von konsensfähigen deskriptiven Analyse-kategorien (und insbesondere die Annotation einer nennenswerten Auswahl von Texten) mühsam und womöglich innerhalb der Literaturwissenschaft mit unzureichendem Renommee³⁰ verbunden – nicht zuletzt angesichts des nur mittelbaren Beitrags zum eigentlichen Erkenntnisinteresse der literarischen Interpretation, wenn man der ‚defensiven‘ Positionierung folgt. Allerdings ist diesem Punkt entgegenzusetzen, dass gerade für die Entwicklung von robusten Computermodellen eine möglichst breite Auseinandersetzung mit dem mühsamen Teil des Unterfangens förderlich ist.³¹ Mittelfristig ist zu hoffen, dass

²⁸Im Zusammenhang einer Diskussion des Stellenwerts der Narratologie (als eine Form der deskriptiven Textanalyse) kommen bspw. auch Kindt und Müller – trotz der Überzeugung, dass sich diese „weder für die Fundierung noch für die Falsifikation oder Evaluation von Interpretationen [eignet]“ – zu dem Schluss, dass sie ein unverzichtbares „heuristisches Potenzial“ für die Interpretation hat, vgl. Tom Kindt/Hans-Harald Müller, „Wieviel Interpretation enthalten Beschreibungen? Überlegungen zu einer umstrittenen Unterscheidung am Beispiel der Narratologie“, in: Jannidis/Lauer/Martínez u. a. (Anm. 19), 286–304, hier: 301.

²⁹So weist Fotis Jannidis in Vorträgen gern darauf hin, dass ein korpusorientiertes Vorgehen in der Praxis der Digitalen Literaturwissenschaft sehr deutlich vor Augen führt, an welchen Stellen Konzepte präzisiert werden müssen, für die konventionell ein weitgehender Konsens angenommen wurde.

³⁰Unglücklich ist insbesondere, dass im Zuge einer systematischen Entwicklung von Kategorien und der Analysepraxis idealerweise bekannte Texte entlang der gängigen Interpretationen ‚deskriptiv durchanalysiert‘ werden müssten, dass gerade dies jedoch mangels Originalität wenig Anerkennung finden dürfte.

³¹Das Stuttgarter Zentrum für reflektierte Textanalyse (CRETA) hat sich daher die interdisziplinäre Methodenentwicklung zur Aufgabe gemacht. Schnell treten Herausforderungen zutage, die sich aus den Anforderungen der textwissenschaftlichen Disziplinen für die komputationelle Analyse über etablierte Verfahren aus der Computerlinguistik und Sprachtechnologie ergeben.

auch im literaturwissenschaftlichen Kernfach die Anerkennung für die zeitaufwändige Methodenentwicklung wächst – und für die damit einhergehenden Veränderungen der Arbeitspraxis (bspw. die fast zwingend erforderliche Teambildung, die zu Publikationen in Co-Autorenschaft führt usw.). Ein Meilenstein dürfte sein, dass ein Beitrag als uneingeschränkt originell anerkannt wird, wenn dieser wesentliche Innovationen in eingesetzten Modellen bzw. Verfahren vorschlägt und experimentell anhand eines literarischen Textkorpus (oder auch eines Einzeltexts) validiert – auch wenn zu keinem literarischen Gegenstand eine neue Fragestellung aufgeworfen bzw. beantwortet wird. Das erhebliche Analysepotenzial, das in der korpusorientierten Praxis auch für literaturhistorische und hermeneutische Arbeiten liegt (das grundsätzlich sicherlich von wenigen infrage gestellt wird), dürfte sich mit einem erweiterten Originalitätsbegriff wesentlich dynamischer entwickeln lassen.

Eine zweite Problematik scheint mir in der Grenzbestimmung für deskriptive Analysekategorien zu liegen: Wo liegt die Schwelle der literaturwissenschaftlichen *Interpretation*, wie weit reicht der Skopus von intersubjektiv reproduzierbaren Kategorien einer professionellen literarischen *Beschreibung* von Texteigenschaften, für die eine weitgehende Unabhängigkeit von Rezeptionskontext und theoretischem Rahmen angesetzt werden kann? Will man sich nicht auf unstrittige Konzepte wie Metadaten zum Text beschränken, begibt man sich schnell auf literaturtheoretisch umstrittenes Terrain.

Eine naheliegende Strategie für das Erreichen intersubjektiver Übereinstimmung in der Annotation wäre, sich grundsätzlich auf formal-strukturelle Texteigenschaften zu beschränken (in der Annahme, diese wären unabhängig von jedem Interpretationsaspekt). Die Diskussion in Abschn. 1.4 (im Zusammenhang mit den Beispielen 1 und 2) hat jedoch gezeigt, dass bereits für unverdächtige strukturelle Kategorisierungen wie zur Satzgliedfunktion ein ‚Durchsickern‘ von Interpretationsunterscheidungen stattfinden kann. Bei der Annotation jegliche Inferenzen aus einem semantischen und pragmatischen Textverständnis auszublenken, scheint also kontraproduktiv.

Aus computerlinguistischer Sicht naheliegend wäre es, im Anschluss an Hirsch³² eine weitestgehend stabile inhaltliche Bedeutungsebene (*Meaning*) anzusetzen – in Abgrenzung von der Bedeutung des literarischen Texts im Sinne von Bedeutsamkeit (*Significance*), welche der literarischen Interpretation vorbehalten bleibt. Wie allerdings Jannidis u. a. historisch-systematisch darlegen,³³ läuft bereits die Festlegung auf eine fixe inhaltliche Textbedeutung den Idealen eines hermeneutischen Ansatzes entgegen. Möglicherweise lässt sich aber das *prinzipielle* Problem aus der praktischen Operationalisierung heraushalten, indem als abstrahierendes Konstrukt bei der Annotation das referenzielle Inhaltsverständnis eines prototypischen Lesers

³²Vgl. Anm. XX.

³³Vgl. Fotis Jannidis/Gerhard Lauer/Matías Martínez u. a., „Der Bedeutungsbegriff in der Literaturwissenschaft. Eine historische und systematische Skizze“, in: Dies. (Anm. 19), 3–32.

bzw. einer prototypischen Leserin³⁴ angesetzt wird (die – in Anlehnung an das linguistische Konstrukt des kompetenten Muttersprachlers/der Muttersprachlerin – all jene Inferenzen ziehen, die aus dem kulturellen Allgemeinwissen und dem aufgebauten textuellen Kontext naheliegend sind). Abgesehen von Sonderfällen (wie vielleicht postmodernen Texten, die bewusst mit naheliegenden Inferenzen spielen) könnte ein solcher Ansatz – unter einer Gegenwarts Perspektive³⁵ auf die Textrezeption – zu hoher intersubjektiver Übereinstimmung führen, ohne für die darunter liegende Ebene der literarischen Interpretation unerwünschte Vorentscheidungen zu treffen.³⁶

Als Illustration mag eine Passage aus Mark Twains *Huckleberry Finn* dienen, in der die Erzählung mit der Ironie spielt, die sich daraus ergibt, dass sich in dieser Szene Huckleberry Finn und Tom Sawyer verkleidet haben: Tom als sein Halbbruder Sid und Huck als Tom. Huck schildert als Ich-Erzähler, wie er verkleidet als Tom in Uncle Silas hineinläuft, nachdem er gerade versucht hat, den Doktor zu Hilfe in ihr Versteck zu holen, weil Tom eine Kugel ins Bein bekommen hat:³⁷

³⁴Inwieweit eine Beschränkung auf die referenzielle Bedeutungsebene effektiv möglich wäre, ist freilich eine Frage, die in der Praxis erprobt werden müsste. Die Vielschichtigkeit der Debatte um geeignete Lesermodelle auf Interpretationsebene lässt vermuten, dass jenseits dieser referenziellen Ebene ein Konsens nur schwer vorstellbar wäre.

³⁵Der Versuch, das Verständnis prototypischer Lesender unter einer diachronischen Herangehensweise historisch zu rekonstruieren, wäre problematisch – ein Ergebnis der Diskussion in der Villa Vigoni. Das Hintergrundwissen muss unvollständig bleiben, und gerade durch die vielfältigen Inferenzen, die in das Textverständnis eingehen, ist es sehr schwierig auszuschließen, dass Annahmen und Konventionen aus der Gegenwarts Perspektive trotz guter Schulung der Annotierenden zu einer Verzerrung führen.

³⁶Hier mag auch ein Blick auf die Praxis der Korpusannotation in der Linguistik und den daraus resultierenden Computermodellierungen ermutigend sein: Die umfassendsten Aktivitäten zur syntaktische Annotation eines englischsprachigen Korpus (der Penn Treebank) fanden 1989–1996 im theoretischen Rahmen der Chomsky'schen Transformationsgrammatik statt, die außerhalb einer Kern-Community hoch umstritten war und bis heute ist. Theoriespezifische Annotationsentscheidungen lassen sich jedoch leicht aus dem Korpus herausfiltern; Analyseentscheidungen, die unabhängig vom theoretischen Rahmen von deskriptivem Wert sind, überwiegen bei Weitem. Und so dient das Korpus bis heute als die Basis für Computermodellierung einer deskriptiven syntaktischen Analyse, sei es in einem theoretisch agnostischen Kontext oder im Rahmen bestimmter Theorien – sogar für vehement konkurrierende Theorien, da eine systematische Übersetzung der Annotationen für die allermeisten Fälle möglich ist.

³⁷In der deutschen Übersetzung von Henny Koch lautet die Passage:

Vom Doktor war weit und breit nichts zu sehen. So renn' ich denn zu seinem Haus und höre, daß er in der Nacht gerufen worden und seitdem nicht wieder heimgekommen sei. Armer Tom, denk' ich, da sieht's böß aus, und setz' mich wieder in Trab, und wie ich um die nächste Ecke biege, renn' ich mit dem Kopf beinah auf Onkel Silas' Magen. Er ruft: ‚Junge, Tom, wo habt ihr denn gesteckt all die Zeit, Bengel, he?‘ ‚Ich – ich hab' gar nicht gesteckt‘, stotter' ich, ‚Sid und ich sind nur immer hinter [Jim] hergewesen.‘ ‚Ja, aber wo denn in aller Welt, wo habt ihr ihn denn gesucht? Eure Tante ist in schöner Angst und Aufregung euret wegen!‘ ‚Das braucht sie gar nicht zu sein‘, sag' ich, ‚uns ist nichts passiert. Wir liefen hinter den Männern und den Hunden drein, [...] und nun ist Sid zur Post, um zu sehen, ob er nichts erfahren könne, und ich wollte eben sehen, ob sich was zu essen auftreiben ließe, und dann wären wir heimgekommen.‘ Wir gingen also zur Post, um nach Sid zu sehen, aber der war natürlich nicht dort.

Siehe Mark Twain, *Huckleberry Finns Abenteuer und Fahrten*, München 1962 [engl., USA, 1885], Kap. 32, verfügbar über gutenberg.spiegel.de.

5. I [...] went for the doctor's house, but they told me he'd gone away [...] and warn't back yet. Well, thinks I, that looks powerful bad for Tom, [...]. So away I shoved, and turned the corner, and nearly rammed my head into Uncle Silas's stomach! He says:
 „Why, Tom! Where you been all this time, you rascal?“
 „I hain't been nowheres,“ I says, „only just hunting for [Jim] – me and Sid.“
 „Why, where ever did you go?“ he says. „Your aunt's been mighty uneasy.“
 „She needn't“ I says, „because we was all right. We followed the men and the dogs [...]. Sid's at the post-office to see what he can hear, and I'm a-branching out to get something to eat for us, and then we're going home.“
 So then we went to the post-office to get „Sid“; but just as I suspicioned, he warn't there [...].³⁸

Die Frage der Referenz der Namen „Tom“ und „Sid“ im Erzählertext vs. in der wörtlichen Rede ist, technisch gesehen, komplex und hängt in hohem Maß von den Inferenzen ab, die auf Basis der inhaltlichen Textbedeutung vollzogen werden müssen. Gleichwohl dürften Leser, die sprachlich in der Lage sind, der Texthandlung zu folgen, das ironische Spiel mit der Referenz ohne Mühe nachvollziehen können – wie auch den Grund für die Verwendung von Anführungszeichen im letzten Satz, nämlich um im Erzählertext an das Referenzsystem des Dialogs anzuschließen. Nichts spricht gegen die Erwartung einer intersubjektiv stabilen Annotation der Zusammenhänge auf unterschiedlichen Erzählebenen bzw. Wahrnehmungswelten, und eine entsprechende Annotation scheint mir für keine denkbare literarische Interpretation vorentscheidend.³⁹

Für die Möglichkeit intersubjektiv stabiler Annotationen von komplexeren Texteigenschaften sprechen auch die positiven Erfahrungen, die das Hamburg-Heidelberger literaturwissenschaftliche Annotationsprojekt *heureCLÉA*⁴⁰ im Umgang mit der Frage der Polyvalenz gesammelt hat: In einem zyklischen Prozess zur Verfeinerung von Annotationsrichtlinien gelang es hier, zu Annotationsentscheidungen, für die der Eindruck nicht zu vermeidender individueller Divergenzen naheliegt, einen intersubjektiven Konsens herbeizuführen, nötigenfalls durch Verwendung von Hilfskategorien.⁴¹

³⁸Die Passage ist Teil von Kap. 41 in der Ausgabe, die sich in gutenberg.org findet: Mark Twain, *Adventures of Huckleberry Finn* (Tom Sawyer's Comrade), New York 1885.

³⁹Die Frage, wie die unterschiedlichen Referenzebenen effektiv annotiert werden können, ist nicht ganz einfach, und in jedem Fall sind derartige Verwechslungsspiele für eine automatische Koreferenz-Analyse sicherlich schwierig. Beides ist jedoch unabhängig von der These, dass das Konstrukt einer stabilen inhaltlichen Textebene für das Gros literarischer Texte eine sinnvolle Abstraktion ist.

⁴⁰*heureCLÉA*, <http://www.heureclea.de> (letzter Aufruf 22.09.2017); eine Kooperation zwischen einer literaturwissenschaftlichen Arbeitsgruppe (unter Leitung von Jan Christoph Meister) und einer Informatikgruppe (geleitet von Michael Gertz).

⁴¹Vgl. Gius/Jacke (Anm. 27).

2.2 Computermodelle in der hermeneutischen Praxis? Die ‚offensive‘ Replik

Obgleich die Vermeidung von unmittelbar interpretationsabhängigen Analysekat­egorien eine gut handhabbare Arbeitspraxis garantiert, kann man die Frage stellen, ob die Anwendungsbedingungen für Computermodelle in den Literaturwissen­schaften wirklich erzwingen, dass die Modelle jeglichen Aspekt der Interpretation – auch innerhalb eines hermeneutisch orientierten Selbstverständnisses – ausblenden. Eine offensivere Replik auf die naheliegenden Einwände erscheint möglich.

Defensiv erscheint zunächst auch Willard McCarty's Grundhaltung, wenn dieser die Modelle der *Digital Humanities* (bzw. des *Humanities Computing*) als für sich jeweils unzulängliche Vehikel auf dem Weg zu einer tieferen Erkenntnis darstellt.⁴² Der Anspruch, mit Modellvarianten eine Empirie vorherzusagen, die unabhängig vom wissenschaftlichen Betrachter besteht, wird negiert. Welche andere Rolle kommt einem Modell aber im Erkenntnisprozess zu? Wir können davon ausgehen, dass sich der interpretierende (digitale) Literaturwissenschaftler oder die Literaturwissenschaftlerin bei der Konzeption des Modells davon überzeugt hat, dass dieses ein gewisses komplexes Zusammenspiel von Texteigen­schaften so gut erfasst, dass es Positivbeispiele für einen Interpretationsansatz von Negativbeispielen besser abgrenzt als andere denkbare Modelle (*Baselines*) – jeweils für einen gegebenen Ausschnitt an Texten und relativ zum für relevant erachteten Kontext. (So mag bspw. für die rezeptionsästhetische Untersuchung eines bestimmten literarischen Kerntexts ein Untersuchungskorpus von Texten zugrunde gelegt werden, die in der nachfolgenden Epoche entstanden sind. Die Korpusanalyse soll klären, ob und in welchen Texten sich die Rezeption des Kern­texts in Form von Textmerkmalen niedergeschlagen hat. Anhand von Texten, zu denen eine Auseinandersetzung der Autorin mit dem Kerntext durch bekannte Fakten nachgewiesen ist, könnte nun ein Modell so eingestellt werden, dass es tendenziell in der Lage ist, zwischen direkten Einflüssen und zufälligen sprachlichen Ähnlichkeiten zu unterscheiden.) Wird nun ein solches Modell auf einige neue Texte angewendet, welchen Status haben dann die Modellvorhersagen? Wenn sie *ausschließlich* als Heuristik für die Auswahl derjenigen Texte dienen, die anschließend einer konventionellen Feinstudie, also dem *Close Reading*, unterzogen würden, so hätten empirische Implikationen des Modells in der Tat keinen Einfluss auf den Kern des Erkenntnisprozesses. Wenn McCarty jedoch von einer sukzessiven Verbesserung der (global zwar unzulänglichen) Vehikel ausgeht, heißt dies doch, dass die Modelle bei einer kontrastiven Betrachtung sehr wohl der *Bewertung* unterschiedlicher Interpretationswege dienen können.⁴³ In die

⁴²Willard McCarty, *Humanities Computing*, London 2005.

⁴³Die Rolle der Computermodelle fügt sich damit ein in die Argumentation aus Jannidis (Anm. 19) zum Umgang mit der These der Polyvalenz von Literatur, in der er darauf verweist, dass die These keinesfalls in Widerspruch zur Zielsetzung stehe, unter den konkurrierenden Interpretationen die besten zu identifizieren (für Jannidis diejenigen, die die „für einen Leser durch die Lektüre eines Textes manifest gewordenen Informationen“ besser als andere erfassen).

Beurteilung, ob ein weiterer analytischer Schritt ein Fortschritt auf dem Weg zu einer tieferen Erkenntnis ist oder nicht, fließt in diesem Fall die Vorhersage des Modells ein.

Der Anspruch an die Validität⁴⁴ eines solchen Modells muss also sein, dass es in den gewählten Parametrisierungen hinsichtlich der aktuell betrachteten Fragestellung mit den interpretatorischen Vorannahmen kompatibel ist und dass die Texte, auf die das Modell angewendet wird, sich in relevanten Eigenschaften in die angenommene Grundgesamtheit einfügen.

Die ‚offensive‘ Replik auf den methodischen Einwand zum Verhältnis zwischen literaturwissenschaftlicher Interpretation und den Ansprüchen einer operationalisierten Annotationspraxis mit entsprechenden Repräsentativitätsannahmen des *Gold-Standard*-Korpus besteht genau in einer Verschiebung des Datenspektrums, auf das die etablierte korpusorientierte Praxis angewendet wird: Der Validitätsanspruch relativ zu einer *intersubjektiv* stabilen Zielkategorisierung als Referenz (aus dem sich empirische Implikationen ableiten lassen) wird ersetzt durch den Anspruch, eine möglicherweise *subjektiv* charakterisierte Zielkategorisierung systematisch zu erfassen und relativ hierzu ein reproduzierbares Vorhersageverhalten auf neuen Daten zu erhalten (für welche eine Zugehörigkeit zur angenommenen Grundgesamtheit unterstellt wird).

Auf diese Weise wird es m. E. möglich, Computermodelle in die Abwägung von textanalytischen Alternativen einzubeziehen, die mit Fragen der literaturwissenschaftlichen Interpretation in Zusammenhang stehen. Zunächst soll das Vorgehen jedoch anhand eines Beispiels konkretisiert werden.

⁴⁴Aus Sicht eines rein hermeneutischen Vorgehens könnte man fragen, weshalb eine Validitätsforderung für Modelle aufgestellt wird, welche im Erkenntnisprozess überwiegend heuristischen Charakter haben sollen – zumal gerne auf die Generierung von neuen Perspektiven nach dem ‚Serendipitätsprinzip‘ verwiesen wird, das heißt, dass die Betrachtung von Vorhersageergebnissen den Blick auf Unerwartetes lenkt und so neue Erkenntnisse befruchtet (vgl. z. B. Dimpel [Anm. 7], 354). Die entscheidende Frage bei den Anforderungen an das Modell scheint mir zu sein, ob ein rein zufälliger Effekt für akzeptabel erachtet würde (wie man ihn ja mit Zufallskomponenten erzeugen könnte und wie er in der digitalen Kunst zum Einsatz kommt). Unter einem wissenschaftlichen Anspruch würde dies sicherlich verneint. Wenn aber ein Modell eine (eventuell noch intuitiv bzw. vorthoretisch gefasste) Regularität erfassen soll, bindet man sich methodisch an die Randbedingungen, nach denen die Modelle operieren; sprich, mindestens implizit wird man bei der Modellanwendung ein Mindestmaß an reproduzierbarem Verhalten entsprechend der eigenen Abstraktionen erwarten. In der Arbeit mit Computermodellen und -werkzeugen bei geisteswissenschaftlichen Gegenständen kommt unabhängig von dieser Überlegung hinzu, dass deren Anwendung in einer Studie (eventuell zu Unrecht) ein Maß an Objektivität suggeriert, das Rechenschaft über die Rahmenbedingungen abzulegen gebietet. Ein nachträgliches Aufdecken von Verletzungen zentraler Modellierungsannahmen kann einen eigentlich vielversprechenden arbeitspraktischen Ansatz in Misskredit bringen.

2.3 Beispiel: Korpusbasierte Modellierung interpretationsgebundener Analysekatogorien

Grundsätzlich wäre denkbar, mit Computermodellen zu arbeiten, die Zielkategorien einer literarischen Interpretation *direkt* modellieren. Dies würde einen recht engen interpretationstheoretischen Rahmen voraussetzen (möglicherweise literatursoziologisch, rezeptionshistorisch etc.), innerhalb dessen Hypothesen zu sinnvollen Interpretationen einiger Zieltexte aufgestellt und mittels eines statistischen Modells überprüft werden, das auf einem mit entsprechenden Interpretationskategorien annotierten Korpus trainiert wurde. Hier erscheint jedoch eine sehr voraussetzungsreiche und zeitaufwendige Annotation unvermeidlich.

Mir erscheint daher zumindest kurzfristig ein Ansatz attraktiver, der weder direkt auf Zielkategorien der literarischen Interpretation zielt noch sich mit intersubjektiv stabilen deskriptiven Analysekatogorien begnügt. Stattdessen bezieht er in der Mitte des Spektrums solche Kategorien der Textanalyse (etwa der narratologischen Textanalyse) ein, die mit interpretatorischen Entscheidungen verwoben sind. Die Etablierung eines studienunabhängigen intersubjektiven *Gold-Standards* ist damit ausgeschlossen, dennoch kann über die unterschiedlichen Textinstanzen bzw. ein ganzes Korpus hinweg ein systematisch ‚reliables‘ Analyseverhalten angestrebt werden, wie es mit einem Computermodell erzielbar ist.

Als Beispielaufgabe soll hier die Frage der subjektiven Perspektivierung von Figuren in Erzählungen dienen, durchgeführt für eine Reihe von Erzähltexten Arthur Schnitzlers in der dritten Person. Darunter sind Texte, in denen die gesamte Erzählung an die Wahrnehmung einer Figur gebunden ist: *Frau Berta Garlan* (1900) und *Casanovas Heimfahrt* (1918) (jeweils konsistent aus der Wahrnehmungswelt der Titelfigur erzählt, wobei die explizite Attribution von Wahrnehmungs-, Gedanken- oder Trauminhalten durch die Erzählerinstanz mit Passagen der erlebten Rede changieren). In der Novelle *Die Toten schweigen* (1897) kontrastiert der zweite Teil mit seiner Innensicht der verheirateten Emma formal deutlich mit dem ersten Teil, der eingangs überwiegend den Blickwinkel von Emmas Liebhaber Franz einnimmt. In dem anschließenden langen Dialog sieht Aurnhammer die Inszenierung einer gemeinsamen Perspektive.⁴⁵ Die Erzählung im Roman *Der Weg ins Freie* (1907) folgt überwiegend der Wahrnehmung des Opernkomponisten Baron Georg von Wergenthin, enthält jedoch auch einige Passagen, die aus der subjektiven Perspektive anderer Figuren geschildert sind.

Charakteristisch für Schnitzlers Erzählstil sind lange Passagen der Innensicht einer Figur, häufig mit umfangreichen Rückblenden, innerhalb derer etwa auch der Hintergrund einer zweiten Figur aus dem subjektiven Blickwinkel der ersten charakterisiert werden kann. Ein Beispiel ist folgende Passage aus dem *Weg ins Freie*, in der die Erzählung Georg von Wergenthins Gedanken folgt und

⁴⁵ Achim Aurnhammer, *Arthur Schnitzlers intertextuelles Erzählen*, Berlin/Boston 2013, 33.

wir – gewissermaßen aus zweiter Hand – eine Schilderung zur Vergangenheit des jüdischen Schriftstellers Heinrich Bermann erhalten, mit dem Georg von Wergenthin befreundet ist. Das subjektive Element der Perspektive Georgs wird erst wieder gegen Ende der Passage deutlich, nachdem zwischenzeitlich die Tatsache leicht in Vergessenheit geraten konnte, dass wir es mit einer „gefärbten“ Schilderung zu tun haben:

6. Georg bummelte langsam gegen die Stadt zu. Er überlegte, ob er ins Kaffeehaus gehen sollte. Er hatte keine rechte Lust dazu. [...] auf Leo Golowskis Kommen war nur selten zu rechnen; und die andern jungen Leute [...] lockten ihn nicht eben an, [...]. Im ganzen fand er den Ton der jungen Leute untereinander bald zu intim, bald zu fremd [...]. [W]ährend er selbst nach wie vor sich ziemlich zurückhaltend verhalten und insbesondere über seine Beziehungen zu Frauen jede Andeutung vermieden, hatte ihm Heinrich nicht nur von der fernen Geliebten erzählt, [...] sondern auch von der Kinder- und Knabenzeit in der kleinen böhmischen Provinzstadt, wo er vor dreißig Jahren zur Welt gekommen war. Sonderbar und zuweilen fast peinlich erschien Georg der wie aus Zärtlichkeit und Widerwillen, aus Gefühlen von Anhänglichkeit und von Losgerissenheit gemischte Ton, in dem Heinrich von den Seinen, insbesondere von dem kranken Vater sprach, der in jener kleinen Stadt Advokat, und eine Zeitlang Reichsratsabgeordneter gewesen war. Ja, er schien sogar ein wenig stolz darauf zu sein, daß er als Zwanzigjähriger schon dem allzu Vertrauensseligen sein Schicksal vorausgesagt hatte, genau so wie es sich später erfüllen sollte: nach einer kurzen Epoche der Beliebtheit und des Erfolgs hatte das Anwachsen der antisemitischen Bewegung ihn aus der deutsch-liberalen Partei gedrängt, die meisten Freunde hatten ihn verlassen und verraten [...]. Heinrich, dem die Phrasen des Vaters von Deutschtum, Freiheit, Fortschritt in all ihrer Ehrlichkeit immer gegen den Strich gegangen waren, hatte dem Niedergang des alternden Mannes anfangs wie mit Schadenfreude zugesehen; allmählich erst [...] stellte bei dem Sohne sich ein verspätetes Mitleid ein. [...] Seine ersten künstlerischen Erfolge fanden in dem verdüsterten Hause der Heimat kein Echo mehr. Dem Vater nahte unter schweren Zeichen der Wahnsinn, und der Mutter [...] versank nun [...] die ganze Welt. [...] Auch von andern Verwandten erzählte Heinrich, deren er aus früherer Zeit sich erinnerte, und ein teils lächerlicher, teils rührender Zug fromm beschränkter alter Juden und Jüdinnen schwebte an Georg vorüber, wie Gestalten einer andern Welt.⁴⁶

Formal sind Schilderungen wie die zu Heinrichs Hintergrund bisweilen ununterscheidbar von denkbaren Einschüben der Erzählerinstanz, welche die interne Fokalisierung des Protagonisten (hier Georg) unterbrechen könnten – eventuell

⁴⁶Die Seitenzählung folgt: Arthur Schnitzler, Gesammelte Werke. Die erzählenden Schriften, 2 Bde., Band 1: Der Weg ins Freie, Frankfurt a. M. 1961, Kap. 2, 706 ff. Eine elektronische Fassung ist verfügbar über zeno.org/.

mag sogar der Eindruck eines Sprungs in der internen Fokalisierung entstehen, hier etwa auf Heinrich. Die Fortführung gegen Ende der zitierten Passage zeigt jedoch (wie in vielen solchen Fällen bei Schnitzler) im Nachhinein, dass wir in der Tat Georgs Wahrnehmung der Schilderungen Heinrichs ihm gegenüber – oder genauer Georgs Erinnerung an diese Schilderungen, die ihm durch den Kopf gehen, während er durch die Stadt bummelt – mit ihm teilen.

Es finden sich immer wieder Passagen (wie folgende Szene in Kap. 3 im Zusammenhang mit einer Radtour), in denen Gedanken und Ansichten einer Figur geschildert werden – hier sind es Gedanken Heinrichs –, für die sich jedoch bei genauerer Betrachtung erweist (oder richtiger: für die aus weitergehenden Betrachtungen die Interpretation nahe liegt), dass dies nicht aus einer wirklichen Innensicht heraus geschieht:

7. Heinrich nickte. [...]

Er versank für eine Weile in Nachdenken, schob sein Rad in leichten, ungeduldigen Stößen vorwärts und war gleich wieder um ein paar Schritte voraus. Dann begann er wieder von seiner Septemberreise zu sprechen. Beinahe mit Ergriffenheit dachte er an sie zurück. Alleinsein, Fremde, Bewegung, war es nicht ein dreifaches Glück, das er genossen? „Was für ein Gefühl von innerer Freiheit mich damals durchfloß“, sagte er, „kann ich Ihnen kaum beschreiben.“ [...].

Georg empfand stets eine gewisse Verlegenheit, wenn Heinrich pathetisch wurde. „Jetzt könnte man vielleicht wieder fahren“, sagte er, und sie schwangen sich auf die Räder.⁴⁷

Vielmehr werden uns die Gedanken Heinrichs aus der Sicht einer anderen Figur vermittelt (hier wieder Georg) – Heinrich hat sie also Georg gegenüber geäußert (selbst wenn der Akt der Äußerung nicht explizit ist – wie hier für den Satz „Beinahe mit Ergriffenheit dachte er an sie zurück.“ und den nachfolgenden Gedanken, bei dem es sich gewissermaßen um erlebte Rede aus zweiter Hand handelt).

Sprachlich finden sich Indikatoren für diese vermittelte Sicht: Zum Inhalt des Nachdenkens Heinrichs (zu Beginn von Passage [7]) erfahren wir zunächst nichts, stattdessen werden Heinrichs Handlungen aus der Außensicht beschrieben (Schieben des Rads), mit deiktischen Bestimmungen relativ zu Georgs Perspektive („ein paar Schritte voraus“). Erst nachdem explizit erwähnt wird, dass Heinrich zu sprechen beginnt, erfahren wir etwas über den Inhalt seiner Gedanken. Nach einer längeren Passage wörtlicher Rede unterstützt die Schilderung von Georgs Verlegenheit wiederum im Nachhinein, dass wir Heinrichs Gedanken von Georgs Warte aus gehört haben.

Die sprachlichen Charakteristika für die mittelbare Fokalisierung Heinrichs (in [7], via Georg) stehen in klarem Kontrast zu Passagen einer internen

⁴⁷ Ebd., 712.

Fokalisierung, für die folgendes Beispiel typisch ist: Die engmaschige Erwähnung von klar perspektivierten Sinneswahrnehmungen fällt auf.

8. Vom Turm der Michaelerkirche schlug es neun, als Georg vor dem Kaffeehaus stand. An einem Fenster, das der Vorhang nicht verhüllte, sah er den Kritiker Rapp sitzen, einen Stoß von Zeitungen vor sich auf dem Tisch. Eben hatte er den Zwicker von der Nase genommen, putzte ihn, und so sah das blasse, sonst so hämisch-kluge Gesicht, mit den stumpfen Augen wie tot aus. Ihm gegenüber, mit ins Leere gehenden Gesten, saß der Dichter Gleißner, im Glanze seiner falschen Eleganz, mir einer ungeheuern, schwarzen Krawatte, darin ein roter Stein funkelte. Als Georg, ohne ihre Stimmen zu hören, nur die Lippen der beiden sich bewegen und ihre Blicke hin- und hergehen sah, faßte er es kaum, wie sie es ertragen konnten in dieser Wolke von Haß sich eine Viertelstunde lang gegenüber zu sitzen.⁴⁸

Schnitzlers Erzählstil eignet sich sehr gut, um die Idee einer teilweise interpretationsgebundenen Modellspezifikation zu illustrieren. Die zuletzt geschilderten Beobachtungen sind in der Zusammenschau plausibel; formal finden sich jedoch keine zwingenden Indikatoren. Im Gegenteil, im Werk einer anderen Autorin oder eines anderen Autors wäre es möglicherweise naheliegend, die Folge „Dann begann er wieder von seiner Septemberreise zu sprechen. Beinahe mit Ergriffenheit dachte er an sie zurück.“ (aus Kap. 7) als interne Fokalisierung zu interpretieren.

Eine detaillierte narratologische Annotation längerer Textpassagen, die den Zusammenhängen gerecht wird, wäre sehr aufwendig. Wahrnehmungsebenen und Erzählebenen müssten mit Fragen der Fokalisierung in Beziehung gesetzt werden. Intersubjektive Übereinstimmung in allen wesentlichen Fällen zu erreichen, scheint zudem unwahrscheinlich, da Einzelinstanzen Interpretationsspielraum lassen (selbst wenn eine einheitliche Gesamttendenz festgelegt werden könnte). Für bestimmte weitergehende Fragen, die Schnitzlers Erzählstil betreffen (etwa zu dessen Entwicklung, zu eventuellen intertextuellen Einflüssen⁴⁹ o.ä.) oder für die interpretationsunterstützende Analyse einzelner Texte könnte jedoch ein ‚flacherer‘ Analyseansatz bereits gewinnbringend sein, der sich auf Grundlage des Leseindrucks recht zügig annotieren lässt.

2.3.1 Kategorisierungsentscheidung und Merkmale für überwachtes Training

Mit dieser Motivation kann mit Blick auf die genannten Schnitzler-Texte eine *Ad-hoc*-Charakterisierung einer interpretationsrelevanten Teilfrage vorgenommen werden: Die Analyseaufgabe wird zugespitzt auf einen studienspezifischen Aspekt der subjektiven Perspektivierung von Figuren in einer heterodiegetischen

⁴⁸ Ebd., 708.

⁴⁹ Hierzu stellt Achim Aurnhammer detaillierte Untersuchungen an (vgl. Anm. 45).

Erzählung. Diese lässt sich – in enger Anlehnung an das Konzept des subjektiven *Point of View*, dessen Fortentwicklung im Verlauf von Erzähltexten Wiebe⁵⁰ mit einem detaillierten algorithmischen Ansatz modelliert – fassen als eine einfache binäre Klassifikationsaufgabe. Jeder textuellen Erwähnung einer Figur (in Kap. 6 etwa „Georg – er – er – [...] Heinrich – er – dem kranken Vater – er – dem allzu Vertrauensseligen“) ist dabei einer von zwei möglichen Zuständen zuzuordnen: (a) die erwähnte Figur wird (an genau dieser Textstelle) als Träger bzw. Trägerin der subjektiven Perspektive verstanden oder (b) dies ist nicht der Fall – d. h. die Erwähnung wird dann entweder verstanden als eingebettet in die subjektive Wahrnehmungswelt oder Binnenerzählung einer anderen Figur, oder als Gegenstand einer externen Fokalisierung oder Nullfokalisierung durch die Erzählerinstanz. Im Folgenden werde ich die Kontexte, die einen Leser zur Annahme von Zustand (a) bewegen,⁵¹ kurz mit dem Begriff ‚interne Fokalisierung‘ bezeichnen – auch wenn im Einzelfall nicht notwendigerweise eine Standarddefinition dieser narratologischen Beschreibungskategorie zutrifft, die auf Genettes⁵² Begriff der Fokalisierung zurückgeht.

Diese sehr zielgerichtete Unterscheidung wird kombiniert mit einer umfassenden computerlinguistischen Analyse der Texte, die für quantitative Studien oder für überwacht maschinelles Lernverfahren reiche Merkmalsrepräsentationen beisteuern. Für die Analyse wurden die englischen Übersetzungen der Texte verwendet.⁵³ In einem ersten Schritt wird wörtliche Rede aus den Texten herausgefiltert (da diese orthogonal zu eventuellen Fokalisierungen im Erzähltext liegt). Der verbleibende Erzählertext wird mit den computerlinguistischen

⁵⁰Janyce M. Wiebe, „Tracking point of view in narrative“, in: *Computational Linguistics* 20/2 (1994), 233–287.

⁵¹Die Unterscheidung lässt sich recht gut durch Texterweiterungstests operationalisieren. Eine Figurenerwähnung wird vom Leser als Träger der subjektiven Perspektive empfunden, wenn beim Lesen ein Zusatz „Er/sie erinnerte sich in diesem Augenblick an [keine] ähnliche(n) Situationen/Ereignisse.“ als bruchlos wahrgenommen würde. Dieser Zusatz könnte bspw. für die Erwähnung von ‚Maria‘ im zweiten Satz der hypothetischen Passage (i) „Karl rannte zu Marias Haus und klopfte an die Tür. Maria machte sofort auf.“ getestet werden. Im Kontext der beiden Sätze in (i) würde „Sie erinnerte sich in diesem Augenblick an eine ähnliche Situation letztes Jahr.“ für die meisten Leser einen unnatürlichen Bruch hervorrufen; naheliegender erscheint es, die subjektive Perspektive bei Karl zu sehen. Anders wäre die Wahrnehmung, wenn es noch vor Passage (i) hieße: „Maria blickte sorgenvoll aus dem Fenster. Plötzlich sah sie eine Gestalt die Straße herauf laufen. Das musste Karl sein.“

⁵²Gérard Genette, *Figures III*, Paris 1972.

⁵³Von gutenberg.org. Die Robustheit der computerlinguistischen Analyse, insbes. für die Koreferenz-Erkennung, ist größer und die eingesetzte stilistische Wortliste ist nur für das Englische verfügbar. Mit der Verwendung von englischen Übersetzungen kann jedoch durchaus die Durchführung von komparatistischen Untersuchungen erleichtert werden. Bspw. konnte auch Gustave Flauberts *Madame Bovary* (1856) in englischer Übersetzung in die Sammlung integriert werden. Aurnhammer (Anm. 45) diskutiert die intertextuellen Bezüge von Schnitzlers *Die Toten schweigen* zu *Madame Bovary*.

Standardwerkzeugen aus der Stanford CoreNLP Suite⁵⁴ analysiert, u. a. auf Ebene von Wortarten und Satzsyntax; Eigennamen werden erkannt, und zuletzt wird eine Koreferenzanalyse vorgenommen, die alle Erwähnungen von referenziellen Ausdrücken (Eigennamen und Pronomina, aber auch definite Nominalphrasen wie „the boy“) zueinander in Beziehung gesetzt, um die Ketten von Ausdrücken zu bestimmen, die auf dieselbe Entität referieren. Die englischen Übersetzungen der Schnitzler-Texte lassen sich recht robust analysieren. Für die Hauptfiguren führt die Koreferenzanalyse zu einem erstaunlich guten Ergebnis.

Wichtige sprachliche Indizien für interne Fokalisierung liegen in den semantischen Klassen der Verben,⁵⁵ als deren Subjekt (oder Objekt) die Figuren fungieren; hinzu kommt die Frage nach Tempus und Aspekt (Rückblenden, also Plusquamperfekt-Passagen, sind ein starker Indikator), ob eine modale Einbettung (etwa unter dem Auxiliar „could“ oder unter „seemed to“) vorliegt und welche adverbialen Bestimmungen Verwendung finden (Negation, Satzadverbien wie „apparently“ etc.).

Um dem hohen Maß an Kontextabhängigkeit der Fokalisierungsinterpretation gerecht zu werden, das oben diskutiert wurde, wird neben dem eigentlichen referenziellen Ausdruck, der zu klassifizieren ist (eine bestimmte Instanz von ‚Georg‘ oder ‚er‘ etc.), ein Fenster von referenziellen Ausdrücken in seinem Vor- und Nachkontext in Betracht gezogen, einschließlich der Information, ob es sich laut automatischer Analyse um einen koreferenten Ausdruck handelt oder nicht. So kann die Tendenz erfasst werden, dass nach einer Verwendung einer Figurenreferenz als Subjekt eines Wahrnehmungsverbs weitere Erwähnungen eher auf interne Fokalisierungen hindeuten.

Zuletzt werden, den Beobachtungen von Brooke u. a. folgend,⁵⁶ lexikalische Stilprofile im Kontext der Figurenausdrücke erhoben. Brooke u. a. konnten in Erzählungen Virginia Woolfs und James Joyces eine fein differenzierte stilistische Unterscheidung der freien indirekten Rede unterschiedlicher Figuren nachweisen. Die sechsdimensionalen Profilvektoren, die sie auf dem gutenbergtag-Korpus induziert haben, stehen innerhalb der *GutenTag*-Werkzeugsuite⁵⁷ zur Verfügung.

All diese Merkmale sind nur Indizien, maschinelle Lernverfahren können jedoch Generalisierungen zu typischen Konstellationen (die teils sicherlich autoren-spezifisch sind) aus den Daten induzieren. Indem neben den oberflächlichen Merkmalen auch relevante grammatische Kategorien bereitgestellt werden, reicht im günstigen Fall bereits eine verhältnismäßig kleine Menge von Beispielen

⁵⁴ Stanford CoreNLP – Natural language software (letzter Aufruf 22.09.2017).

⁵⁵ Hier kann für das Englische die umfassende Verbtaxonomie in VerbNet genutzt werden, mit Zugriff über den Unified Verb Index, <https://verbs.colorado.edu/verb-index/> (letzter Aufruf 22.09.2017).

⁵⁶ Vgl. Anm. 5.

⁵⁷ GutenTag, <http://www.cs.toronto.edu/~jbrooke/gutentag/> (letzter Aufruf 22.09.2017); darin `built_in_lexicons/sixstyleplus.txt`. Ich danke Julian Brooke für diesen Hinweis.

dafür aus, die stärksten Muster aufzugreifen (die komplexen und subtilen Interaktionen zwischen Sprach- und Textebenen würden sonst eine sehr große Menge von relativ homogenen Trainingsdaten erforderlich machen).

2.3.2 Annotation und Analysemöglichkeiten

Da die eigentliche Analyseaufgabe auf eine partiell interpretationsabhängige binäre Entscheidung reduziert wurde (hinter der selbstverständlich komplexe Zusammenhänge stehen, die in jedem Einzelfall durch eine komplexere narratologische Analyse tiefer zu durchdringen wären), ist eine sehr zügige Annotation von Beispieldaten möglich. In wenigen Stunden konnten so über 1000 Einzelinstanzen von Personenreferenzen innerhalb der Figurenrede annotiert werden (als (a) intern oder (b) nicht intern fokalisiert).

Im Rahmen dieser experimentellen Studie wurde bewusst ein subjektiver Annotationsansatz gewählt: Ich habe die Textpassagen unter der Hypothese gelesen, dass Schnitzler eine interne Fokalisierung forciert und wir auch in langen Analepsen jeweils die übergeordnete Innensicht der Figur erleben, die die Passage eröffnet. Wo immer sprachliche Mittel *erzwingen*, dass die Innensicht zu einer anderen Figur wandert, wurde dies konsequent annotiert – Instanzen jedoch, die sprachlich einen Interpretationsspielraum boten, wurden entsprechend der Hypothese markiert.⁵⁸

Das Vorgehen ist also zwar subjektiv, aber systematisch. Es bedient sich des in Abschn. 1.1 angeführten Gedankens einer bewussten idealisierenden Annahme: Obleich ohne Zweifel weitere narratologisch relevante Faktoren im Spiel sind, wird der Blick auf eine für zentral gehaltene Dichotomie gelenkt (das Experiment ignoriert gewissermaßen das Pendant der Reibung in einem Experiment in der Mechanik, da laut Annahmen andere Effekte dominieren).

Die idealisierende Abstraktion lässt sich nicht rein empirisch verteidigen. Sie wird stets aus Überlegungen motiviert sein, die mit einer Interpretationshypothese zusammenhängen – idealerweise speist sie sich aus einer bereits verschiedentlich abgesicherten Theorie und einem etablierten Konsens über sinnvolle Abstraktionen. Indem jedoch hypothesengesteuerte Annotationen an realen Textdaten vorgenommen werden, wird es möglich, Implikationen dieser Hypothesen empirisch auf Korpusdaten hin zu überprüfen.⁵⁹

⁵⁸An dieser Stelle sollte deutlich werden, dass eine andere Annotationsstrategie möglich gewesen wären – bspw. hätte man in Beispiel (7), das ‚eingebettet‘ erlebte Rede einer weiteren Figur enthält, diesen Fokalisierungsaspekt annotieren können.

⁵⁹Das Beispiel der Annotation einer narratologischen Kategorie (Fokalisierung/subjektive Figurenperspektive), deren Zuweisung relativ stark mit dem Leseverständnis von impliziten Textmerkmalen zusammenhängt, habe ich aus der Überlegung heraus gewählt, dass hier das oben diskutierte ‚Durchsickern‘ von weitergehenden interpretatorischen Annahmen sicherlich deutlich zutage tritt. Gleichwohl unterstützt die Modellierung ‚nur‘ eine narratologische Textanalyse, es wird hier keine Hypothese zur Textinterpretation insgesamt modelliert. Diskussionen in der Villa Vigoni und im Nachgang haben mir vor Augen geführt, dass aus literaturwissenschaftlicher Sicht durch den narratologischen Charakter meines Beispiels das beabsichtigte ‚offensive‘ Vordringen in Bereiche der hermeneutischen Textinterpretation für manchen nach wie vor nur eingeschränkt

2.4 Perspektiven zum Einsatz von interpretatorisch gefärbten Vorhersagemodellen

Mit der Annotation erschließen sich eine Reihe von Szenarien für den Einsatz von Computermodellen: Auf den handannotierten Beispielen kann zunächst mit überwachten Lernverfahren ein Klassifikator trainiert werden. Durch Anwendung auf nicht annotierte Textpassagen können – evtl. auf nach Figuren vorgefilterten Daten – rasch Bereiche ausfindig gemacht werden, für die bestimmte Fokalisierungsmuster vorhergesagt werden.

Beispielsweise kann ein Vorhersagemodell, das auf Passagen aus dem *Der Weg ins Freie* (für die Figuren Georg und Heinrich) sowie aus *Frau Berta Garlan* trainiert wurde, auf Passagen im *Der Weg ins Freie* angewandt werden, in denen Georgs Geliebte Anna Rosner präsent ist. Bei der Lektüre wird schnell deutlich, dass im gesamten Roman der Erzähler Annas Innensicht so gut wie nie einnimmt.⁶⁰

Zu den Passagen, für die das Modell eine Häufung von vermuteter interner Fokalisierung ausgibt, gehört (9a); die Passage (9b) ist ein Beispiel für eine geringe Wahrscheinlichkeit für interne Fokalisierung.

9. a) She had for the first time in her life the infallible feeling that there was a man in the world who could do anything he liked with her.⁶¹
- b) Anna had given herself to him without indicating by a word, a look or gesture that so far as she was concerned, what was practically a new chapter in her life was now beginning.⁶²

exemplifiziert wird. Man hätte sicherlich die Modellierung von wesentlich weitergehenden Hypothesen zur Textinterpretation experimentell angehen können. Andererseits dürfte deutlich werden, dass bei der Textanalyse ein Kontinuum der Interpretationsabhängigkeit besteht. Sofern es gelingt, eine literaturwissenschaftliche Praxis zu etablieren, die auch in ‚tieferen‘ Bereichen dieses Kontinuums Annotationsverfahren und Computermodelle gewinnbringend einsetzt, würde ich dies für einen wichtigen Durchbruch erachten. Das sukzessive Erschließen von mehr und mehr interpretationsabhängigen Analysekatoren erscheint zudem aus Sicht der systematischen Operationalisierung und des Korpusaufbau strategisch sinnvoll.

⁶⁰So ist auf der Wikipedia-Seite zum *Der Weg ins Freie* notiert (ohne Belege): „Schnitzler erlaubt das Denken nur Georg, dessen Freund Heinrich in Ausnahmefällen und Anna fast gar nicht.“ (Wikipedia. Die freie Enzyklopädie, https://de.wikipedia.org/wiki/Der_Weg_ins_Freie [letzter Aufruf 22.09.2017]).

⁶¹Die Passage befindet sich am Ende von Kap. 2: Schnitzler (Anm. 46), 710: „[...] zum erstenmal in ihrem Leben mit dem untrüglichen Gefühl, daß es einen Menschen auf der Welt gab, der aus ihr machen konnte, was ihm beliebte.“

⁶²Anfang des dritten Kapitels, vgl. Schnitzler (Anm. 46), 711: „Anna hatte sich ihm gegeben, ohne mit einem Wort, einem Blick, einer Gebärde anzudeuten, daß nun für sie gewissermaßen ein neues Kapitel ihres Lebens anfang.“

Tab. 1 Experimente zum überwachten Training von verschiedenen Klassifikatoren auf unterschiedlichen Konstellationen von Trainings- und Testdaten

		(i) Logistic Regression				(ii) Gaussian Naive Bayes			
		Precision	Recall	F-Score	Accuracy	Precision	Recall	F-Score	Accuracy
(A) Training: Weg, Test auf Garlan	Intern fok.	0,77	0,71	0,74	0,78	0,63	0,65	0,64	0,67
	Nicht intern fok.	0,79	0,84	0,82		0,70	0,68	0,69	
(B) Training: Weg, Test auf Weg	Intern fok.	0,73	0,77	0,75	0,75	0,75	0,78	0,77	0,74
	Nicht intern fok.	0,76	0,73	0,74		0,74	0,70	0,71	
(C) Training: Weg + Garlan, Test auf Weg	Intern fok.	0,76	0,73	0,75	0,79	0,57	0,54	0,55	0,63
	Nicht intern fok.	0,81	0,83	0,82		0,67	0,69	0,68	

Neben solchen sehr ermutigenden Instanzen kommt es selbstverständlich auch zu wenig nachvollziehbaren Vorhersagen – die Klassifikationsaufgabe ist sehr subtil, sodass bei der eher kleinen Zahl von Trainingsdaten noch kein sehr robustes Ergebnis erzielt wird. Als ein Filter könnte die Vorhersage jedoch einen Explorationsprozess unterstützen.

Von größerem Belang für das hier vorgeschlagene Vorgehen sind allerdings systematische Untersuchungen, welche die Vorhersagen der Modelle gar nicht für eine automatische Analyse von Texten einsetzen. Stattdessen kann versucht werden, aus dem Generalisierungsverhalten der Modelle empirisch fundierte Rückschlüsse zu ziehen auf die Tragfähigkeit der eigenen Hypothesen zur Konzeptualisierung einer Analyseidee (hier zur internen Fokalisierung bei Schnitzler).

Tab. 1 zeigt beispielsweise den Vergleich von Trainingsexperimenten mit unterschiedlichen Kombinationen von Trainings- und Testdaten (A bis C), wobei zudem zwei unterschiedliche maschinelle Lernverfahren zur Anwendung kommen ([i] Logistic Regression und [ii] *Gaussian Naive Bayes*).⁶³ Für jedes Experiment werden Evaluationsergebnisse auf manuell annotierten Testdaten gelistet, nach den beiden Zielklassen (,intern fokalisiert‘ vs. ,nicht intern fokalisiert‘) aufgeteilt, mit *Precision* (Treffgenauigkeit der Vorhersage), *Recall* (Fähigkeit, alle Instanzen zu finden), einem Mittelwert daraus (*F-Score*) sowie der Genauigkeit der Gesamtvorhersage (*Accuracy*). Eine *Precision* von 0,77 des Modells für interne Fokalisierung in Szenario (A)/(i) sagt uns beispielsweise, dass es bei 77 % der Figurenerwähnungen, für die das Modell als positive Instanzen vorhergesagt

⁶³ Für die eigentlichen Lernexperimente wurde die Python-Bibliothek scikit-learn verwendet (<http://scikit-learn.org> [letzter Aufruf 22.09.2017]), die systematische Vergleiche dieser Art relativ einfach macht.

hat, eine Übereinstimmung mit der manuellen Vergleichsannotation gab; der *Recall* von 0,71 besagt, dass 71 % der manuell als intern fokalisiertem Instanzen auch vom Modell als solche erkannt wurden. Der *Accuracy*-Wert 0,78 sagt aus, dass von allen Vorhersagen (also den positiven und den negativen Fällen) 78 % in der gleichen Kategorie landen wie die manuelle Annotation.

Das Szenario A suggeriert, dass die Modellierung der Aufgabe einen relativ effektiven Grad der Abstraktion erreicht: Das Modell, das auf Beispielen in Schnitzlers *Der Weg ins Freie* trainiert wurde, erzielt auch auf bei der Anwendung auf Textinstanzen aus *Frau Berta Garlan* relativ ordentliche Vorhersageergebnisse: *Accuracy* 0,78 (allerdings gelingt dies mit dem *Gaussian-Naive-Bayes*-Ansatz (ii) wesentlich schlechter – 0,67 – als mit *Logistic Regression* (i)). Die Ergebnisse in Szenario B (in dem die Menge der nutzbaren Trainingsdaten geringer ist als in A, da die Testdaten ausgeklammert bleiben) ist beim Verfahren i) etwas schwächer als bei A; beim Verfahren ii) verhält es sich umgekehrt.

In Szenario C werden ‚gemischte‘ Trainingsdaten aus den beiden Texten verwendet, und mit Verfahren (i) kann so eine Verbesserung der Vorhersagegenauigkeit gegenüber B erreicht werden (*Accuracy* 0,79 gegenüber 0,75). Dies deutet auf eine verhältnismäßige Homogenität der Datenlage hin. Allerdings zeigt die Abhängigkeit von der Wahl des Verfahrens (für Verfahren (ii) ist die Mischung der Daten nachteilig), dass derartige Schlussfolgerungen einer zusätzlichen Absicherung bedürfen.

Um zu beurteilen, welche sprachlichen Merkmale bei der Induktion des Klassifikatorverhaltens eine Rolle spielen, können Vergleichsexperimente durchgeführt werden, in denen bestimmte Lernmerkmale ausblendet werden (ein sogenanntes *Feature-Ablation*-Experiment). Tab. 2 zeigt eine Wiederholung der hellgrünen Zellen aus Tab. 1, die mit einem Experiment kontrastiert werden (rötlich), in dem die verfügbaren Merkmale reduziert waren: Die oben erwähnten Merkmale, die auf lexikalischen und strukturellen Analysen basieren, tragen also – zumindest für dieses Szenario – zur Generalisierungskapazität bei.⁶⁴

Zuletzt sei eine naheliegende Erweiterung des Analyseskopus erwähnt, die hier ebenfalls nur in Form einer knappen Pilotstudie angedeutet werden kann, die aber von großem Wert für hypothesengeleitete Untersuchungen zu interpretationsrelevanten Eigenschaften sein dürfte: Die bisher gezeigten Analysen basierten auf manuell annotierten Textinstanzen. Trotz der Möglichkeit eines verhältnismäßig zügigen Fortschreitens bei der Annotation bleibt damit die erreichbare Datenmenge beschränkt (oder es muss ein großer Annotationsaufwand betrieben werden).

Wenn wir jedoch von Texten ausgehen, die narratologisch einem sehr homogenen Muster folgen – wie im Untersuchungskorpus *Frau Berta Garlan* und *Casanovas Heimfahrt* –, so könnten wir hier die idealisierende Annahme ansetzen,

⁶⁴Dies ist in der Tat nicht in allen Szenarien der Fall. Eine differenziertere Untersuchung wäre nötig, um ein genaueres Verständnis zur Rolle der unterschiedlichen Merkmalsklassen zu entwickeln.

Tab. 2 Experimente zur Rolle der Merkmale beim Lernen (*Feature-Ablation*)

			Logistic Regression			
			Precision	Recall	F-Score	Accuracy
Training mit gesamte Merkmalsmenge	(C) Training: <i>Weg</i> + <i>Garlan</i> , Test auf <i>Weg</i>	Intern fok.	0,76	0,73	0,75	0,79
		Nicht intern fok.	0,81	0,83	0,82	
Training ohne linguistische Merkmale	(C) Training: <i>Weg</i> + <i>Garlan</i> , Test auf <i>Weg</i>	Intern fok.	0,61	0,68	0,64	0,71
		Nicht intern fok.	0,78	0,73	0,75	

dass schlichtweg alle Vorkommen der jeweiligen Titelfigur in die Kategorie ‚intern fokalisiert‘ fallen – und entsprechend jede andere Figur in die Kategorie ‚nicht intern fokalisiert‘. Auf Basis der automatischen Koreferenzresolution lässt sich so mit sehr geringem Aufwand eine große Menge von quasi-annotierten Trainingsdaten erzeugen. Dies kann einerseits aus Sicht der Vorhersageoptimierung von Interesse sein – es lassen sich möglicherweise robustere Vorhersagemodelle trainieren, da viele relevante Muster erst mit größeren Datenmengen ins Blickfeld wandern. Im Kontext dieses Beitrags ergibt sich jedoch wiederum eine literaturanalytische Perspektive, auf die ich gleich zu sprechen komme.

Zunächst sei das Verfahren anhand des Beispielkorpus illustriert. Tab. 3 wiederholt erneut die Ergebnisse für Szenario C aus Tab. 1. Daneben wird nun ein neues Szenario (D) gestellt, bei dem die Zahl der Trainingsinstanzen von 1.034 auf 3.476 erhöht wird, indem für jeweils ein Kapitel aus *Frau Berta Garlan* und *Casanovas Heimfahrt* eine automatische Annotation nach dem geschilderten Schema hinzugefügt wurde.

Der Vergleich der Analysequalität zeigt in diesem Fall, dass die *Accuracy* konstant bleibt – sprich, die zusätzlichen Daten schaden nicht der mittleren Vorhersagegenauigkeit, erhöhen sie aber auch nicht. Die klassenspezifische Evaluation zeigt eine Verschiebung im Vorhersageverhalten: Die nicht fokalisierten Instanzen werden genauer erkannt (*Precision*), und mehr Zielinstanzen dieser Art werden entdeckt (*Recall*); dies jedoch auf Kosten des *Recalls* bei den intern fokalisierten Instanzen. Dieses Verhalten überrascht wenig, wenn man sich klar macht, dass die synthetischen Daten wesentlich mehr Instanzen vom Typ ‚nicht fokalisiert‘ enthalten. Ein feingliedrigeres Vorgehen ist für die Zukunft jedoch ohne Weiteres denkbar.

Tab. 3 Experimente zur Verwendung von ‚quasi-annotierten‘ Trainingsdaten

		Logistic Regression			
		Precision	Recall	F-Score	Accuracy
(C) Training: <i>Weg + Garlan,</i> Test auf <i>Weg</i>	Intern fok.	0,76	0,73	0,75	0,79
	Nicht intern fok.	0,81	0,83	0,82	
(D) Training: <i>Weg + ‚Auto‘-Garlan</i> <i>+ ‚Auto‘-Casanova,</i> Test auf <i>Weg</i>	Intern fok.	0,73	0,57	0,64	0,79
	Nicht intern fok.	0,82	0,90	0,86	

Nun abschließend zu den interpretationsrelevanten Anwendungsmöglichkeiten der (interpretatorisch motivierten) synthetischen Annotation: Abseits einer möglichen Verbesserung der Robustheit dürften die zuletzt diskutierten Möglichkeiten eines hypothesengesteuerten Textvergleichs an Attraktivität gewinnen, wenn für sie nicht jeder Text mit einer manuellen Referenzannotation versehen werden muss. Sofern das Modellverhalten anhand einiger sorgfältig annotierter Referenzdatensätze kalibriert wurde, erscheint es durchaus denkbar, dass dann eine kontrastive Textanalyse verschiedener Vergleichsgegenstände unter idealisierenden Annahmen zur Homogenität gemacht werden.⁶⁵

Das Vorgehen bei der Erstellung einer quasi-manuellen Annotation muss dabei nicht der Intuition des oder der Experimentierenden folgen, sondern kann auch Thesen aus der Forschungsliteratur aufnehmen und gewissermaßen auf den Prüfstand stellen. So lässt sich für *Die Toten schweigen* mit geringem Aufwand eine synthetische Referenzannotation erstellen, die der oben angeführten Analyse

⁶⁵Der hier propagierte strategische Einsatz von idealisierenden Annahmen, die den typischen Fall verallgemeinern und damit einen (vorläufigen) systematischeren Zugang zur Empirie erschließen sollen, kann (muss eventuell sogar) zu einem Zwischenstadium in der Modellbildung führen, das aus der Warte eines historisch-hermeneutischen Vorgehens den Vorwurf der ‚Unterkomplexität‘ auf sich ziehen könnte. Da jedoch Computermodelle für einen seriösen Einsatz laufend in ihrer technischen Parametrisierung kalibriert werden müssen, scheint mir ein solches Vorgehen unumgänglich. Für eine systematische Weiterentwicklung der formal-komputationellen Modellierungsoptionen schiene ein Baukasten von für sich gesehen jeweils ‚unterkomplexen‘ Theiemodulen die besten Voraussetzungen zu bieten.

Aurnhammers⁶⁶ folgt und den zweiten Teil komplett aus der Perspektive Emmas fokalisiert.

Nicht zuletzt ließen sich auch intertextuelle Bezüge grundsätzlich mit dem angedeuteten Verfahren in die Betrachtung einbeziehen. Aurnhammer verweist beispielsweise auf die Parallelen zwischen *Die Toten schweigen* und Gustave Flauberts *Madame Bovary*. Letzterer Text konnte mit wenig Aufwand in das Untersuchungskorpus aufgenommen werden (für komparatistische Untersuchungen von Texten der Weltliteratur kann sich der Weg über die englische Übersetzung in gutenberg.org gelegentlich als hilfreich erweisen). Allerdings scheint mir bei der Ausweitung des Korpus auch Vorsicht geboten: Die überraschend ‚glatten‘ Übergänge bei der Arbeit mit den Schnitzler-Texten sind sicherlich dem hohen Grad an stilistischer Homogenität zu verdanken, die allzu eklatante Verletzungen der Repräsentativitätsannahmen in den statistischen Modellierungsverfahren verhindern. Es wird eine Daueraufgabe bleiben, die Bedingungen jeder Verallgemeinerung von Modellen und Methoden auf weitere Gegenstände kritisch zu reflektieren und forschungsstrategische Idealisierungen regelmäßig zu überprüfen.

3 Schluss

Dieser Artikel ging aus von Überlegungen zu den zumeist impliziten Annahmen über die Anwendungsbedingungen von prädiktiven Computermodellen für Aspekte der Textanalyse: Ein Modell oder Werkzeug, das bei der Überprüfung manuell annotierter Testdaten eine ordentliche Vorhersagequalität erreicht, kann im tatsächlichen Anwendungskontext dann als belastbar betrachtet werden, wenn i) Test- und Anwendungstexte als Stichprobe aus derselben Grundgesamtheit betrachtet werden können und ii) die Zielannotation so operationalisiert ist, dass auf dieser Grundgesamtheit intersubjektiv stabile Ergebnisse erzielt werden. Ein methodologisch reflektiertes Vorgehen muss entsprechend kritisch mit der studienspezifischen Datenlage umgehen, und gerade bei literaturwissenschaftlich anspruchsvollen Untersuchungen kann nicht davon ausgegangen werden, dass verfügbare Werkzeuge bzw. verwandte Korpusressourcen ohne größeren Aufwand für die eigenen Analyseziele genutzt werden können. Ein unreflektierter Werkzeugeinsatz kann leicht zu unzulässigen Schlussfolgerungen führen.

In der Community der *Digitalen Literaturwissenschaft* besteht allerdings ein Problembewusstsein, und es wird weitgehend eine Strategie verfolgt, die ich als ‚defensive‘ Reaktion auf die Methodenproblematik bezeichnet habe: Korpusannotationen und Analysemodelle werden vordringlich für solche Analysefragen vorangetrieben, für die verlässliche Ergebnisse erzielt werden können, für die jedoch die empirische Basis erst breiter erschlossen werden muss. Fragen, die von

⁶⁶Vgl. Anm. 45.

unmittelbarer Relevanz für literarische Interpretation sind, werden damit bewusst tendenziell zurückgestellt, da einem methodisch sauberen Vorgehen – zurecht – der Vorrang eingeräumt wird.

In den letzten Abschnitten dieses Beitrags habe ich in Ergänzung zum defensiven Vorgehen ‚offensivere‘ Einsatzformen von korpusbasierten Computermodellen diskutiert – ohne damit suggerieren zu wollen, dass der Weg über die sorgfältige Ausarbeitung von umfassend annotierten Textkorpora ersetzt werden könnte oder sollte. Um innerhalb einer anspruchsvollen Textanalyse empirisch gestützte Argumente zu einem längeren Text oder einem gesamten (Teil-)Korpus anzuführen, deren Analysekategorien nicht interpretationsunabhängig operationalisierbar sind, können prädiktive Computermodelle eine wichtige Funktion erfüllen: Anhand einiger charakteristischer Texte bzw. Textausschnitte können Trainingsinstanzen für eine (partiell subjektive) Analysekatgorie bestimmt werden. Ein Computermodell, das auf den entsprechenden Daten trainiert wird und (günstigenfalls⁶⁷) eventuelle Ebenen übergreifende Muster erfassen kann, lässt sich anschließend auf andere Texte und Textausschnitte anwenden. Die entstehenden modellbasierten Analyseergebnisse können beispielsweise mit einer interpretierenden Lektüre des Zieltextes durch professionelle Lesende verglichen werden und erlauben so vielfältige Rückschlüsse über die Natur eventueller Textunterschiede.

Die Anwendung eines trainierten Modells dürfte dabei in den wenigsten Fällen dem eigentlichen Selbstzweck einer automatischen Kategorisierungsvorhersage folgen – gewissermaßen als Ersatz für die eigene Lektüre. Die Vorhersagegenauigkeit wird angesichts der subtilen Unterscheidungen kaum ein Niveau erreichen, das hierfür befriedigend wäre. Modellierungsexperimente und der Abgleich mit manuellen Annotation könnten aber durchaus einen Platz in der Praxis der Entwicklung von literaturwissenschaftlichen Konzepten und Theorien finden: Mit dem Vergleich unterschiedlicher Modelle hinsichtlich ihrer Annotationsvorhersagen kann die Adäquatheit bzw. der heuristische Wert verschiedener theoretischer Konstrukte (und deren Operationalisierung im Rahmen der Textanalyse) überprüft werden. Und dies in einer Weise, die einen sehr problematischen *Bias* des bzw. der Betrachtenden umgeht, indem die systematischen Aspekte einer interpretationsabhängigen Analyseentscheidung im Computermodell gekapselt werden.

Danksagung Die Arbeiten zu diesem Beitrag wurden teilweise vom deutschen *Bundesministerium für Bildung und Forschung (BMBF)* im Rahmen des Projekts *CRETA (Centrum für reflektierte Textanalyse)* gefördert. Ich bin sehr dankbar dafür, dass ich am *DFG-Symposium „Digitale Literaturwissenschaft“* 2017 in der *Villa Vigoni* teilnehmen konnte, und danke der *DFG* und den Organisatoren herzlich dafür. Ein herzlicher Dank geht auch an alle Beteiligten

⁶⁷ Um das skizzierte Vorgehen in einer methodologisch reflektierten Form umzusetzen, müssen sicherlich noch Evaluationsverfahren ausgearbeitet werden, mit denen sich die Experimentatoren ein präziseres Bild von der Verlässlichkeit der trainierten Modelle verschaffen können.

für die Diskussionen. Besonders danken für Kommentare, Hinweise und Anregungen zum Methodendiskurs, den dieser Artikel berührt, möchte ich Janina Jacke, Fotis Jannidis, Jan Christoph Meister, Axel Pichler, Nils Reiter und Marcus Willand.

Literatur

- Sämtliche digitalen Referenzen wurden letztmalig am 22.09.2017 eingesehen.
- Aurnhammer, Achim, *Arthur Schnitzlers intertextuelles Erzählen*, Berlin/Boston 2013.
- Bamman, David/Underwood, Ted/Smith, Noah A., „A Bayesian Mixed Effects Model of Literary Character“, in: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 370–379, Baltimore, MD, 2014.
- Bresnan, Joan, *Lexical-Functional Syntax*, Oxford 2001.
- Brooke, Julian/Hammond, Adam/Hirst, Graeme, *Using Models of Lexical Style to Quantify Free Indirect Discourse in Modernist Fiction. Digital Scholarship in the Humanities*, 2016, DOI <https://doi.org/10.1093/lc/fqv072>.
- Burrows, John, „Delta‘. A Measure of Stylistic Difference and a Guide to Likely Authorship“, in: *Literary and Linguistic Computing* 17 (2002), 267–287.
- Dimpel, Friedrich M., „Der Computerphilologe als Interpret – ein Teilzeit-Empiriker?“, in: Jan Borkowski/Stefan Descher/Felicitas Ferder u. a., *Literatur interpretieren. Interdisziplinäre Beiträge zur Theorie und Praxis*, Münster 2015, 339–359.
- Elson, David K./McKeown, Kathleen R., „Automatic attribution of quoted speech in literary narrative“, in: *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI'10)*, Palo Alto 2010, 1013–1019.
- Elson, David K./Dames, Nicholas/R. McKeown, Kathleen, „Extracting social networks from literary fiction“, in: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, ACL '10*, Stroudsburg, PA, 2010, 138–147.
- Genette, Gérard, *Figures III*, Paris 1972.
- Gius, Evelyn/Jacke, Janina, *Zur Annotation narratologischer Kategorien der Zeit. Guidelines zur Nutzung des CATMA-Tagsets*, Version 2, Hamburg, November 2016.
- Hage, Volker, „Mein Name sei Ovid. Ein großer Roman. Christoph Ransmayrs ‚Die letzte Welt‘“, in: *DIE ZEIT* 41 (1988), <http://www.zeit.de/1988/41/mein-name-sei-ovid/>.
- Hammond, Adam/Brooke, Julian/Hirst, Graeme, „A tale of two cultures: bringing literary analysis and computational linguistics together“, in: *Proceedings of the NAACL 13 Workshop on Computational Linguistics for Literature*, Atlanta, GA, 2013, 1–8.
- Hirsch, Eric D. Jr., *Validity in Interpretation*, Yale 1967.
- Hovy, Eduard/Lavid, Julia, „Towards a ‚science‘ of corpus annotation. A new methodological challenge for corpus linguistics“, in: *International Journal of Translation* 22/1 (2010), 13–36.
- Jannidis, Fotis/Lauer, Gerhard/Martínez, Matías u. a., „Der Bedeutungsbegriff in der Literaturwissenschaft. Eine historische und systematische Skizze“, in: Dies. (Hg.), *Regeln der Bedeutung. Zur Theorie der Bedeutung literarischer Texte*, Berlin/New York 2003, 3–32.
- Jannidis, Fotis, „Polyvalenz – Konvention – Autonomie“, in: Ders./Gerhard Lauer/Matías Martínez u. a. (Hg.), *Regeln der Bedeutung. Zur Theorie der Bedeutung literarischer Texte*, Berlin/New York 2003, 305–328.
- Jannidis, Fotis/Lauer, Gerhard, „Burrows’s Delta and Its Use in German Literary History“, in: Matt Erlin/Lynne Tatlock (Hg.), *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*, Rochester 2014, 29–54.
- Jannidis, Fotis/Krug, Markus/Reger, Isabella u. a., *Automatische Erkennung von Figuren in deutschsprachigen Romanen. Digital Humanities im deutschsprachigen Raum (DHd)*, Graz 2015, URN nbn:de:bvb:20-opus-143332.
- Jannidis, Fotis/Lauer, Gerhard/Martínez, Matías u. a. (Hg.), *Regeln der Bedeutung. Zur Theorie der Bedeutung literarischer Texte*, Berlin/New York 2003.

- Kao, Justine/Jurafsky, Daniel, „A Computational Analysis of Style, Affect, and Imagery in Contemporary Poetry“, in: *Proceedings of the Workshop on Computational Linguistics for Literature (Conference of the North American Chapter of the Association for Computational Linguistics. Human Language Technologies, NAACL-HLT)*, Montréal 2012, 8–17.
- Kindt, Tom/Müller, Hans-Harald, „Wieviel Interpretation enthalten Beschreibungen? Überlegungen zu einer umstrittenen Unterscheidung am Beispiel der Narratologie“, in: Fotis Jannidis/Gerhard Lauer/Matías Martínez u. a., *Regeln der Bedeutung. Zur Theorie der Bedeutung literarischer Texte*, Berlin/New York 2003, 286–304.
- Kuhn, Jonas: „Computerlinguistische Textanalyse in der Literaturwissenschaft? – oder: ‚The Importance of Being Earnest‘ bei quantitativen Untersuchungen“, in: Andrea Albrecht/Sandra Richter/Marcel Lepper u. a., *Quantitative Verfahren in der Literaturwissenschaft. Von einer Scientia Quantitatis zu den Digital Humanities*, Berlin/Boston 2018, 11–44.
- Kuhn, Jonas, „Computational text analysis within the Humanities. How to combine working practices from the contributing fields?“, in: *Language Resources and Evaluation* (June 2019), DOI <https://doi.org/10.1007/s10579-019-09459-3>.
- Kuhn, Jonas/Blessing, André, „Die Exploration biographischer Textsammlungen mit computerlinguistischen Werkzeugen – methodische Überlegungen zur Übertragung komplexer Analyseketten in den Digital Humanities“, in: Ágoston Z. Bernád/Christine Gruber/Maximilian Kaiser (Hg.), *Europa baut auf Biographien. Aspekte, Bausteine, Normen und Standards für eine europäische Biographik*, Wien 2018, 225–257.
- McCarty, Willard, *Humanities Computing*, London 2005.
- Moretti, Franco, „Network Theory, Plot Analysis“, in: *Stanford Literary Lab Pamphlets 2* (Mai 2011), <http://litlab.stanford.edu/LiteraryLabPamphlet2.pdf>.
- Ransmayr, Christoph, *Die letzte Welt, Roman, mit einem Ovidischen Repertoire*, Nördlingen 1988.
- Schnitzler, Arthur, *Gesammelte Werke. Die erzählenden Schriften*, 2 Bde., Band 1: *Der Weg ins Freie*, Frankfurt a. M. 1961.
- Trilcke, Peer, „Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft“, in: Philip Ajouri/Katja Mellmann/Christoph Rauen (Hg.), *Empirie in der Literaturwissenschaft*, Münster 2013, 201–247.
- Twain, Mark, *Adventures of Huckleberry Finn (Tom Sawyer’s Comrade)*, New York 1885.
- Twain, Mark, *Huckleberry Finns Abenteuer und Fahrten*, München 1962 [engl., USA, 1885].
- Vala, Hardik/Jurgens, David/Piper, Andrew u. a., „Mr. Bennet, his coachman, and the Archbishop walk into a bar but only one of them gets recognized. On the difficulty of detecting characters in literary texts“, in: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, hg. von Association for Computational Linguistics, Lissabon 2015.
- Weitin, Thomas, „Digitale Literaturwissenschaft“, in: *Deutsche Vierteljahresschrift für Literaturwissenschaft und Geistesgeschichte* 89/4 (2015), Sonderheft: *Zur Lage der Literaturwissenschaft. Aktuelle Bestandsaufnahmen und Perspektiven*, 651–656.
- Wiebe, Janyce M., „Tracking point of view in narrative“, in: *Computational Linguistics* 20/2 (1994), 233–287.
- Willand, Marcus, „Hermeneutische Interpretation und digitale Analyse. Eine Verhältnisbestimmung“, in: Luisa Banki/Michael Scheffel (Hg.), *Lektüren. Positionen zeitgenössischer Philologie*, Trier 2017, 77–100.

Online-Ressourcen

- Wikipedia, *Verwechslungswitz*, <http://de.wikipedia.org/wiki/Verwechslungswitz>.
- Stanford CoreNLP – Natural language software, <https://stanfordnlp.github.io/CoreNLP/>.
- Unified Verb Index, <https://verbs.colorado.edu/verb-index/>.
- GutenTag, www.cs.toronto.edu/~jbrooke/gutentag/.

Wikipedia. Die freie Enzyklopädie, Der Weg ins Freie, https://de.wikipedia.org/wiki/Der_Weg_ins_Freie.

scikit-learn, <http://scikit-learn.org>.

Project Gutenberg, <http://www.gutenberg.org/>.

Stanford CoreNLP – Natural language software, <https://stanfordnlp.github.io/CoreNLP/>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.



Was verändert sich eigentlich? Korpusanalytisch basierte Wissenschaftsgeschichte der germanistischen Praxis am Beispiel der *Deutschen Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte*

Fotis Jannidis, Steffen Martus, Leonard Konle und Jörn Kreutel

1 Einleitung

Im Folgenden fragen wir nach dem Stellenwert von ‚Digitalität‘ für die Geschichte der literaturwissenschaftlichen Germanistik. Es geht uns theoretisch, methodisch und methodologisch um die Frage, wie digitale Daten und Datenverarbeitungsverfahren den Blick auf die Fachgeschichte verändern, welche

Leonard Konle hat die PDF-Scans der *DVjs* mittels OCR bearbeitet und das Ergebnis in ein strukturiertes Format überführt; außerdem hat er die bibliographischen Referenzen aus den Fußnoten der *DVjs*-Texte extrahiert. Jörn Kreutel verantwortet den Abschnitt zur Vernetzung der Zeitschriften und die dahinterstehende Analyse der *Germanistik* in den Jahren 1960–1980. Fotis Jannidis und Steffen Martus sind für die übrigen Ausführungen verantwortlich.

F. Jannidis (✉)
Universität Würzburg, Würzburg, Deutschland
E-Mail: fotis.jannidis@uni-wuerzburg.de

S. Martus
Humboldt-Universität zu Berlin, Berlin, Deutschland
E-Mail: steffen.martus@hu-berlin.de

L. Konle
Frankfurt am Main, Deutschland

J. Kreutel
Berliner Hochschule für Technik, Berlin, Deutschland
E-Mail: joern.kreutel@beuth-hochschule.de

Untersuchungsrichtungen insbesondere durch makroskopische Verfahren nahegelegt werden und welche Erkenntnisinteressen dabei eher in den Hintergrund treten. Die wissenschaftspolitische und normative Komponente des Projekts liegt darin, dass wir uns eine Beschreibung des Fachs wünschen, die die Vielfalt der Funktionen und Leistungen angemessen in den Blick rückt und so etwa auch besser bestimmen kann, was sich durch das Auftreten von *Digital Humanities* verändert, wo Anschlussprobleme bestehen und Kooperationsmöglichkeiten verbessert werden könnten.

Die Selbstbeschreibung der deutschen Philologie und Literaturwissenschaft arbeitet traditionell mit starken Homogenitätsunterstellungen. Dies hat nicht zuletzt mit der Autosuggestion durch den schwer vermeidbaren Kollektivsingular ‚die Germanistik‘ zu tun, der seinen Gegenstand in der Regel schon deswegen verfehlt, weil – je nach Standpunkt – eher die Sprachwissenschaft gemeint ist oder eine Institution, die auf sehr verschiedene Lehrämter vorbereitet, oder – wie im Folgenden – die Ältere oder Neuere deutsche Literaturwissenschaft (ÄdL bzw. NdL). Aber selbst innerhalb der einzelnen Teilbereiche zeigt sich schnell, wie problematisch und unfruchtbar es ist, in einem hohen Maß ‚Einheit‘ vorauszusetzen. Entweder werden dann einzelne Perspektiven zu allgemeinen Diagnosen über den Zustand des Fachs, seine Aufgaben und seine Zukunft hochgerechnet, und man ist leicht enttäuscht, wenn die Idee des Fachs mit dessen Realitäten kollidiert, oder man kapituliert vor der Vielfalt, konstatiert die zunehmende Unübersichtlichkeit und den Verlust an Einheit, wobei natürlich auch dabei unterstellt wird, das Fach könne insgesamt überblickt werden.¹ Wie aber gelingt eine angemessene und differenzierte Beschreibung einer ganzen Disziplin oder eines Fachs? Es liegt nahe, sich wichtige Impulse von jenen digitalen Verfahren zu versprechen, die mit größeren Datenmengen arbeiten.

¹Beide Effekte lassen sich in Debatten um den Zustand der Germanistik beobachten, etwa in der durch den Artikel von Martin Doerry ausgelösten Diskussion: „Schiller war Komponist“. 80.000 junge Menschen studieren Germanistik, das beliebteste geisteswissenschaftliche Fach an den deutschen Hochschulen. Ihre Berufsaussichten sind jedoch ungewiss, ihre Professoren spielen in der Öffentlichkeit keine Rolle“, in: *Der Spiegel* 6 (2017), 104–109; vgl. dazu: Steffen Martus, „Germanistik in der Krise? Der eierlegende Wollmilchgermanist wird dringend gesucht. Die deutsche Literaturwissenschaft taugt nichts, meint der ‚Spiegel‘, doch diese Kritik läuft ins Leere, in: *Frankfurter Allgemeine Zeitung* 33 (8. Februar 2017), 9, <https://www.faz.net/aktuell/feuilleton/forschung-und-lehre/germanistik-in-der-krise-der-eierlegende-wollmilchgermanist-wird-dringend-gesucht-14865806.html>; Heinz Drügh/Susanne Komfort-Hein/Albrecht Koschorke, „Wir Todgeweihten grüßen euch“. Der ‚Spiegel‘ ruft die Krise der Germanistik aus. Und schiebt uns dreien die Schuld zu. Hier bekennen wir, was wir tun“, in: *Frankfurter Allgemeine Zeitung* 34 (9. Februar 2017), 11, <https://www.faz.net/aktuell/feuilleton/debatten/krise-der-germanistik-antwort-von-heinz-druegh-susanne-komfort-hein-und-albrecht-koschorke-14868192.html>; zuletzt sei verwiesen auf Eva Geulen, „Für die Einzelsprachlichkeit der Literatur. Nebenbemerkungen zum jüngsten Streit um die Germanistik“, in: *ZfL Blog* (17. Februar 2017), <http://www.zflprojekte.de/zfl-blog/2017/02/17/eva-geulen-fuer-die-einzelsprachlichkeit-der-literatur-nebenbemerkung-zum-juengsten-streit-um-die-germanistik/> (letzter Aufruf der Links 21.09.2017).

In beiden Varianten des Krisendiskurses dominiert die doxographische Auffassung des Fachs, das bevorzugt aus Perspektive von Theorien, Programmatiken und darauf bezogenen Selbstbeschreibungen rekonstruiert wird. Demgegenüber werden die wissenschaftlichen Routinen unterschätzt, die hinter dem Rücken der Akteure ablaufen, den Alltag strukturieren und für eine robuste Praxis sorgen.² Nimmt man hingegen eine praxeologische Haltung ein,³ verlieren zum einen Theorien und Programme ihren privilegierten Status, weil ‚Theoretisieren‘ und ‚Programmieren‘ dann nur noch Praktiken neben anderen sind und gefragt werden kann, wie solche Operationen in eine komplexe Praxis eingebunden werden.⁴ Zum anderen erweist sich ein Fach wie ‚die Germanistik‘ dann – so eine der zu überprüfenden Ausgangshypothesen – auf eine regelmäßige und gleichmäßige Weise als heterogen; es setzt sich aus Sets von Praktiken zusammen, die durch Familienähnlichkeiten zusammengehalten werden; diese Ensembles von Praktiken bilden „Arbeitseinheiten“⁵ aus und lassen bemerkenswert große, aber nicht beliebig viele Spielräume.

So gesehen, erscheint der Zustand der literaturwissenschaftlichen Germanistik insofern weniger krisenhaft, als viele jener Charakteristika, die als problematisch wahrgenommen werden, keine fachspezifischen Ausnahmererscheinungen sind. Somit stellt sich eher die Frage, in welchen Fächern die strukturelle Normalität Besorgnis erregt und wann sie keine Probleme bereitet. Zudem unterbreitet die Praxeologie alternative Einheitsangebote. Die ‚Einheit des Fachs‘ liegt demnach weniger in dem auf Differenz und Unterschied bedachten deklarativen Wissen als vielmehr im prozeduralen Wissen, nicht im „knowing that“, sondern im „know how“:⁶ in den alltäglichen Verrichtungen, in den eingeübten Aktivitäten und in jenen Routinen, die zu vertraut und unscheinbar sind, als dass sie bedenkenswert erscheinen. So hat etwa die programmatische und theoretische Problematisierung der Autor-Kategorie sehr viel mehr Aufmerksamkeit auf sich gezogen als deren erstaunlich ungebrochenes

²Lorraine Daston, „Die unerschütterliche Praxis“, in: Rainer M. Kiesow/Dieter Simon (Hg.), *Auf der Suche nach der verlorenen Wahrheit. Zum Grundlagenstreit in der Geschichtswissenschaft*, Frankfurt a. M./New York 2000, 13–25.

³Steffen Martus/Carlos Spoerhase, „Praxeologie der Literaturwissenschaft“, in: *Geschichte der Germanistik* 35/36 (2009), 89–96; vgl. nun ausführlich: dies., *Geistesarbeit. Eine Praxeologie der Geisteswissenschaften*, Berlin 2022.

⁴Vgl. Gilbert Ryle, *Der Begriff des Geistes*, Stuttgart 1992, 28, sowie mit Blick auf die Literaturwissenschaft: Steffen Martus, „Wandernde Praktiken ‚after theory‘?“, in: *Internationales Archiv für Sozialgeschichte der deutschen Literatur* 40 (2015), 177–195.

⁵Hans-Jörg Rheinberger, *Experimentalsysteme und epistemische Dinge. Eine Geschichte der Proteinsynthese im Reagenzglas*, Frankfurt a. M. 2006, hier: z. B. 9; Steffen Martus, „Epistemische Dinge der Literaturwissenschaft?“, in: Andrea Albrecht/Lutz Danneberg/Olav Krämer u. a. (Hg.), *Theorien, Methoden und Praktiken des Interpretierens*, Berlin/Boston 2015, 23–51, hier: 49 ff.

⁶So die Unterscheidung von Gilbert Ryle (Anm. 6).

Weiterleben in Interpretationsverfahren;⁷ und die geläufige Privilegierung eines ‚weiten‘ gegenüber einem ‚engen‘ Literaturbegriff hat die normativen Gewohnheiten (z. B. der Erwartung besonderer epistemischer Leistungen von ‚Literatur‘) ebenso wenig wie die Verfahren der literaturhistorischen Kodifizierung tiefgreifend verändert⁸ – im Gegenteil: Es lässt sich gerade unter der Bedingung einer digitalen Erweiterung von Untersuchungskorpora ein eher konservatives, konventionelles Aufmerksamkeits- und Selektionsverhalten feststellen.⁹

Für die Frage nach einer *Digitalen Literaturwissenschaft* ist dieser Befund in zweierlei Hinsicht relevant: Zum einen stellt sich die Frage nach den Quellen der Praxis,¹⁰ mithin danach, inwiefern digitale Analyseverfahren das literaturwissenschaftliche Alltagsgeschäft erhellen. Konkret: Inwiefern sind die Ausweitungen und die (derzeit) *damit einhergehenden* Limitierungen des Untersuchungsdesigns durch korpusanalytische Verfahren für eine praxeologische Beobachtung der Germanistik in besonderer Weise aufschlussreich? Eine gewisse analytische Flachheit, die aus Perspektive einer nicht immer gut informierten Kritik an den *Digital Humanities* grundsätzlich behauptet wird,¹¹ könnte sich nicht nur als Vorteil erweisen, weil damit die enorme Ausweitung des Blicks erkaufte wird, sondern weil dies auch zur Untersuchung von Aspekten zwingt, die trivial erscheinen und gerade deswegen einen ungewohnten Blick auf die Selbstverständlichkeiten eines Fach erlauben. Zum anderen sollte sich die Titelfrage unseres Beitrags („Was verändert sich eigentlich?“) gerade auch im Blick auf die Digitalisierung stellen. Wie also lassen sich entsprechende Veränderungen beschreiben, wenn das prozedurale Wissen fokussiert und der Kollektivsingular ‚die Germanistik‘ ebenso vorsichtig verwendet würde wie ‚die Digitalisierung‘?

⁷ Simone Winko, „Autor-Funktionen. Zur argumentativen Verwendung von Autorkonzepten in der gegenwärtigen literaturwissenschaftlichen Interpretationspraxis“, in: Heinrich von Detering (Hg.), *Autorschaft. Positionen und Revisionen*, Stuttgart/Weimar 2002, 334–354; Marcus Willand, „Autorfunktionen in literaturwissenschaftlicher Theorie und interpretativer Praxis. Eine Gegenüberstellung“, in: *Journal of Literary Theory* 5 (2011), 279–302. Auch im Blick auf die Kategorie des ‚Werks‘ scheint zu gelten, dass diese allenfalls in die Theorie ‚zurückkehren‘ kann, nicht aber in die Interpretationspraxis, aus der sie nicht verschwunden war: Carlos Spoerhase, „Was ist ein Werk? Über philologische Werkfunktionen“, in: *Scientia Poetica* 11 (2007), 276–344, hier: 277.

⁸ Simone Winko/Fotis Jannidis/Gerhard Lauer, „Geschichte und Emphase. Zur Theorie und Praxis des erweiterten Literaturbegriffs“, in: Jörn Gottschalk/Tilmann Köppe (Hg.), *Was ist Literatur? Basistexte Literaturtheorie*, Paderborn 2006, 123–154.

⁹ So z. B. im Blick auf den Stellenwert von Goethes *Wahlverwandtschaften*: Fotis Jannidis/Gerhard Lauer, „Burrow’s Delta and Its Use in German Literary History“, in: Matt Erlin/Lynne Tatlock (Hg.), *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*, Rochester 2014, 29–54.

¹⁰ Vgl. den Schwerpunkt: „Historische Praxeologie. Quellen zur Geschichte philologischer Praxisformen. 1800–2000“, In: *Zeitschrift für Germanistik* 2 (2013), passim.

¹¹ So der ethnographische Befund bei: Smiljana Antonijević, *Amongst Digital Humanists. An Ethnographic Study of Digital Knowledge Production*, Basingstoke/New York 2015, hier: z. B. 47.

Als Experimentierfläche haben wir die *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* (*DVjs*) ausgewählt.¹² Dies hat zunächst einen pragmatischen Grund: Es liegt eine stichprobenartige quantitative Analyse der Zeitschrift für den Zeitraum von 1960 bis 2009 vor, die die oben genannten Hypothesen stützt,¹³ nämlich den Befund einer stabilen und normalen Heterogenität. Der vorliegende Beitrag kann einige dieser Befunde als Vorlage nutzen und sie mit Verfahren der digitalen Analyse in einer Serie von ‚Experimenten‘ überprüfen – allein dieses Vorgehen ist bereits charakteristisch für die Einstellung vieler *Digital Humanists*.¹⁴ Für unsere Korpusanalyse haben wir die *DVjs* folgendermaßen aufbereitet: Die Texte wurden aus dem PDF-Format in TEI-Dokumente konvertiert, wobei vor allem einige wichtige Strukturelemente (Überschrift, Absatz, Fußnoten) ausgezeichnet wurden.¹⁵ Außerdem wurden innerhalb der Fußnoten die einzelnen bibliographischen Referenzen mit einem regelbasierten Verfahren extrahiert. Insgesamt handelt es sich um rund 2234 Dateien, die Aufsätze enthalten, welche in der *DVjs* von 1923 bis 2009 erschienen sind. Das Korpus besteht aus rund 25.686.069 Tokens, also Worten und Satzzeichen.¹⁶

Neben den pragmatischen Gründen unserer Experimentalanordnung gibt es eine Reihe von Eigenschaften, die die *DVjs* als geeigneten Analysegegenstand auszeichnen: Die Zeitschrift reagiert auf eine Phase der Fachgeschichte, in der der Methodenpluralismus offensiv reflektiert wird, sodass konzeptionell neue Formen der Disparität akzeptabel erschienen. Die *DVjs* reagierte darauf affirmativ: Sie erkannte die „verschiedenen Richtungen der Literaturgeschichte“ an, und zwar als „lebendiges Nebeneinander“, das für die „heutige Lage dieser Wissenschaft charakteristisch“ sei. Zwar wurden Editionsprojekte, Miszellen, „alles Reinbiographische“ und ‚bloße‘ „Materialsammlungen“ ebenso ausgeschlossen wie „rein stoffliche Quellenuntersuchungen“, gleichwohl galt der

¹²Da die Texte der *DVjs* zum größten Teil durch das Urheberrecht geschützt sind, können wir unser Untersuchungskorpus leider nicht zur Verfügung stellen. Deshalb ist auch die Weitergabe der Analyse-Skripte in der Form von *Jupyter*-Notebooks nicht hilfreich. Allerdings sind die hier verwendeten Verfahren mit Ausnahme der Netzwerkanalysen so schlicht, dass sie einfach nachvollziehbar sind.

¹³Steffen Martus/Erika Thomalla/Daniel Zimmer, „Die Normalität der Krise. Beobachtungen zur Geschichte der deutschen Literaturwissenschaft aus Fußnotenperspektive“, in: *DVjs* 89 (2015), 510–520.

¹⁴Zu einer praxeologischen Auffassung des Felds: Antonijević (Anm. 13); Friederike Schruhl, „Literaturwissenschaftliche Wissensproduktion unter dem Einfluss der Digitalisierung“, in: *Zeitschrift für Germanistik* N. F. 17 (2017), 37–58. Nach Fertigstellung unseres Aufsatzes ist erschienen: Friederike Schruhl, *Formationen der Praxis. Studien zu Darstellungsformen von Literaturwissenschaft und Digital Humanities*, Göttingen 2020.

¹⁵Vgl. <http://www.tei-c.org> (letzter Aufruf 21.09.2017).

¹⁶Wir können zwar die Anzahl der gezählten Dateien und Tokens genau angeben, jedoch haben sich beim automatischen Zerlegen der Dateien Fehler eingeschlichen. Aufgrund der OCR-Mängel und anderer Fehlerquellen gilt dies auch für die Anzahl der Tokens.

Grundsatz, dass „philologische Strenge und Gewissenhaftigkeit selbstverständliche Voraussetzung bleiben müssen“.¹⁷ Mit anderen Worten: Die *DVjs* entwickelte ihr Profil durch einen produktiven Umgang mit Spielräumen (etwa zwischen ‚positivistischer‘ Mikrologie und ‚geistesgeschichtlichen‘ Syntheseversprechen, die in den 1920er Jahren in die Krise geraten waren).¹⁸ Sie eignet sich daher gut für die Untersuchung der Einheit oder Disparität von Literaturwissenschaft. Dies gilt gerade auch deswegen, weil die Zeitschrift zwar einen eindeutigen Schwerpunkt im germanistischen Bereich hat, sich aber dezidiert als interdisziplinäres Unternehmen versteht und sich in unterschiedlichen Graden um Internationalität bemüht.

Dass die *DVjs* immer mehr gewesen ist als eine „Zeitschrift für deutsche Philologie“, stellt eine technische Herausforderung dar: Sie enthält von Anfang an und bis heute neben Texten aus den Bereichen älterer und neuerer deutscher Literaturwissenschaft auch Beiträge aus anderen Philologien, z. B. der Anglistik und der Romanistik, aber auch Studien aus anderen Nachbardisziplinen, z. B. aus der Musikwissenschaft und vor allem in den ersten Jahrzehnten aus der Philosophie. Seit den 1960er Jahren enthält sie auch Beiträge in englischer Sprache. Aus der Perspektive einer quantitativen Analyse sind dies Störvariablen, insbesondere weil keine Metadaten existieren, die es erlauben, die Fach-Gruppen systematisch zu kontrollieren. Daher war es notwendig, in einem ersten Schritt diese Metadaten automatisch zu erheben, was aber nicht in der feinen Granularität möglich war, die wünschenswert wäre. Zuerst wurden die Texte nach ihrer Sprache unterteilt. Von den mehr als 2200 Beiträgen sind rund 2040 in deutscher Sprache verfasst und 185 in englischer Sprache.¹⁹ Um die Fächer zu erkennen, haben wir rund 200 Texte manuell als mediävistisch, neuphilologisch oder sonstig gelabelt, allerdings sind die Klassen sehr ungleichgewichtig vertreten: neuphilologisch: 127, mediävistisch: 46, sonstig: 35. Grundlage für die Erkennung, die sogenannten

¹⁷So die beiden Herausgeber Paul Kluckhohn und Erich Rothacker im „Vorwort“, in: *DVjs* 1 (1923), V; vgl. dazu: Holger Dainat/Rainer Kolk, „Das Forum der Geistesgeschichte“, in: Robert Harsch-Niemeyer (Hg.), *Beiträge zur Methodengeschichte der neueren Philologien. Zum 125jährigen Bestehen des Max Niemeyer Verlags*, Tübingen 1995, 111–134, hier: insb. 111–117. Zur Entstehung des Vorworts und der Arbeit am Zeitschriftenprogramm: Christoph König, „Individualität, Autonomie, Originalität. Zur Rezeption Diltheys in den ersten Jahren der *Deutschen Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte*“, in: *DVjs* 67 (1993), 197–220, hier: insb. 206–210.

¹⁸Zur Problematisierung der Kampf Begriffe ‚Positivismus‘ und ‚Geistesgeschichte‘ vgl. Tom Kindt/Hans-Harald Müller, „Dilthey gegen Scherer. Geistesgeschichte contra Positivismus. Zur Revision eines wissenschaftshistorischen Stereotyps“, in: *DVjs* 74 (2000), 685–709; zur Problematisierung der Geistesgeschichte in der *DVjs* und in deren Umfeld: Dainat/Kolk (Anm. 19), 122–127.

¹⁹Hierzu wurde die Spracherkennungsbibliothek *langdetect* verwendet, eine Pythonportierung von Googles Spracherkennungsalgorithmus; vgl. <https://pypi.python.org/pypi/langdetect> (letzter Aufruf 21.09.2017). Grundlage waren die ersten 2000 Zeichen, was bei einigen wenigen Texten zu Fehlern geführt hat, z. B. weil sie mit einem sehr langen spanischen Zitat begonnen haben oder weitgehend aus einer Bibliographie bestehen, die Einträge in mehreren Sprachen enthält.

Features, waren alle Wörter. Dennoch ist es erstaunlich, dass die *Accuracy* der Erkennung recht niedrig liegt, nämlich durchschnittlich nur bei 75 %.²¹ Für die im Folgenden diskutierten Untersuchungen wurden sämtliche deutschsprachigen oder die deutschsprachigen neuphilologischen Texte verwendet.

Ein zweiter Aspekt, der die *DVjs* als Gegenstand auszeichnet, liegt in der programmatischen Kontinuität, die jedoch nur in bestimmten Hinsichten auch als pragmatische Beständigkeit gedeutet werden kann. Tatsächlich agiert die Zeitschrift auf eine relativ stabile Weise als Teil des Sozialsystems ‚Wissenschaft‘ und lässt sich von außen wenig irritieren. Dies verhindert einerseits nicht, dass sich Anschlüsse zum Nationalsozialismus ergeben, weil die Germanistik ohnehin in großen Teilen Affinitäten zu dessen Gedankengut und Vokabular aufwies.²² Andererseits mussten bei der Fortsetzung der *DVjs* in der Nachkriegszeit nach der erzwungenen Pause in den Jahren 1945 bis 1948 erneut relativ wenige Anpassungen vorgenommen werden. Die Periode zwischen 1933 und 1945 markierte konzeptionell kaum eine Zäsur, auch wenn kosmetische Anpassungsleistungen erbracht wurden und Erich Rothacker, einer der beiden ersten Hauptherausgeber, ein begeisterter Nationalsozialist war, der antisemitische Ausgrenzungen befördert hat.²³ In einem Gutachten der *Reichsstelle zur Förderung des deutschen Schrifttums* jedenfalls wurde 1938 bemängelt, dass sich nach Lektüre der *DVjs* nicht der Eindruck einstelle, 1933 habe sich „in Deutschland [...] manches geändert“, vielmehr erscheint es dem ‚Amt Rosenberg‘ so, als verschließe sich das Periodikum „bewusst und hermetisch gegen jeden Hauch

²¹ Mit einer sehr großen Streuung von 60 % bis 80 %. Verwendet wurde eine lineare *Support Vector Machine* mit *Stochastic Gradient Descent* aus der Python-Bibliothek *scikit-learn*: <https://scikit-learn.org> (letzter Aufruf 21.09.2017). Nach einigen Tests erwies sich das übliche Gewichtungsverfahren für inhaltliche Dokumentklassifikation, *tf-idf*, auch hier als besser. Die fünffache *Cross-validation* zeigte die erwähnte große Streuung. Eine nicht-technische Einführung zum maschinellen Lernen bietet Ethem Alpaydin, *Machine Learning. The new AI*, Cambridge, MA/London 2016. Erwähnt sei noch, dass *Clustering*-Verfahren auf der Grundlage der ersten hundert häufigsten Wörter keine Gruppenbildungen erbracht haben. *Clustering*-Verfahren verwenden keine Metadaten, sondern zeigen evtl. vorhandene Gruppen auf der Grundlage der verwendeten Merkmale.

²² Holger Dainat, „Anpassungsprobleme einer nationalen Wissenschaft. Die Neuere deutsche Literaturwissenschaft in der NS-Zeit“, in: Petra Boden/Holger Dainat (Hg., unter Mitarbeit von Ursula Menzel), *Atta Troll tanzt noch. Selbstbesichtigungen der literaturwissenschaftlichen Germanistik im 20. Jahrhundert*, Berlin 1997, S. 103–126, S. 108 f.; ders., „wir müssen ja trotzdem weiter arbeiten“. Die *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* vor und nach 1945“, in: Wilfried Barner/Christoph König (Hg.), *Zeitenwechsel. Germanistische Literaturwissenschaft vor und nach 1945. Eine Veröffentlichung der Arbeitsstelle für die Erforschung der Geschichte der Germanistik im Deutschen Literaturarchiv Marbach am Neckar*, Frankfurt a. M. 1996, 76–100; zur *DVjs*: Elke Dubbels, „Zum Verhältnis von wissenschaftlicher Tradition und Politik im ‚Dritten Reich‘. Die *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* in den Jahren 1933–1944“, in: *DVjs* 78 (2004), 672–706, hier: 696 u. 699 f.

²³ Dainat/Kolk (Anm. 19), 129–134; Dubbels (Anm. 23); Dainat (Anm. 23), 77–80 u. 91 f.

der Gegenwart“.²⁴ Ähnliches galt für die Kontinuität über die politische Grenze des Jahres 1945 hinweg.²⁵ Entsprechend bestätigte ein Verlagsprospekt zur Fortsetzung der *DVjs* im Jahr 1949 die Programmatik von 1923, ergänzt um die Wertschätzung für die bereits vor 1945 vertretenen und von der Werkimmanenz geadelten „Interpretationen“, mit einem neuen Akzent auf komparatistischen Interessen sowie mit einer verringerten Reserve gegenüber „Funden“.²⁶

Diese additive programmatische Ergänzung lässt erwarten, dass sich das Spektrum lediglich erweiterte und die Herausgeber ein noch umfangreicheres Integrationsangebot als zuvor anboten. Tatsächlich aber verbergen sich etwa hinter dem Stichwort ‚Interpretation‘ neue Schreibweisen; die Art der in der *DVjs* veröffentlichten Texte veränderte sich offensichtlich, und es fällt leicht, einen Beitrag von 1923 von einem Aufsatz des Jahres 2003 zu unterscheiden. Wie also bildet sich das programmatische Selbstverständnis in jenen Befunden ab, die sich aus einer Makroperspektive ergeben? Wie entwickelt sich die Zeitschrift unter der Federführung unterschiedlicher Herausgeber? Und wie korreliert die Entwicklung der Zeitschrift mit der Fachentwicklung? Verändern sich die Darstellungsformen bei gleichbleibend breitem Spektrum von Zugängen zur ‚Literatur‘, sodass die *DVjs* Einblicke in den Zustand der literaturwissenschaftlichen Germanistik insgesamt gewährt?

Diesem letzten Aspekt haben wir uns für den Zeitraum von 1960 bis 1980 über eine Netzwerkanalyse genähert, die mit den Daten der *Germanistik* operiert.²⁷ Geht man davon aus, dass „Fachzeitschriften“ das wissenschaftliche Feld strukturieren, „indem sie Fächer, Forschungsschwerpunkte und Gruppenbildungen sichtbar machen [...]“;²⁸ dann stellt sich die Frage, wie man diese Strukturen identifiziert – angesichts der sehr großen Datenmengen, die Zeitschriften produzieren: Gilt die *DVjs* zurecht als „führende[] literaturwissenschaftliche[] Zeitschrift Deutschlands“?²⁹ Ein Blick in das Verlagsarchiv der *DVjs* im *Deutschen Literaturarchiv*, Marbach a.N., unterstützt diese Einschätzung insofern, als die Herausgeber aus einem beträchtlichen Überangebot auswählen können und stets mehr Ab- als Zusagen verteilen. Dieser Befund zeigt zudem, dass den Zeitschriftenmachern erhebliche Gestaltungsmöglichkeiten zur Verfügung standen.

²⁴Dainat/Kolk (Anm. 19), 132.

²⁵Dies zeigt sich am ersten Aufsatz des Heftes, das nach der Zwangspause erscheint: Werner Günthers Aufsatz „Über die absolute Poesie“ wurde von den Herausgebern bereits 1944 diskutiert, eröffnete dann programmatisch den ansonsten kommentarlosen 23. Jahrgang und legte ein Zeugnis dafür ab, dass die Arbeit an der Zeitschrift im Prinzip nie aufgehört hatte: Dainat (Anm. 23), 80 f.

²⁶Dainat (Anm. 23), 86–90.

²⁷Wir sind dem Verlag *Walter de Gruyter*, namentlich Manuela Gerlof, zu großem Dank verpflichtet, da uns die Datenbank zu Forschungszwecken zur Verfügung gestellt wurde.

²⁸Ebd., 76.

²⁹Ebd.

Die Analyse von Zeitschriften- und Kopublikationsnetzwerken auf Grundlage der *Germanistik*, des „Internationalen Referatenorgans mit bibliographischen Hinweisen“, bestätigt den Stellenwert der *DVjs*.³⁰ Dabei muss einkalkuliert werden, dass sich das bedeutende Informationsmedium, wie die Dokumente des Redaktionsarchivs (*DLA*, Marbach a.N.) zeigen, vergeblich um Vollständigkeit bemüht – die Verantwortlichen machen sich darüber keine Illusionen.³¹ Gleichwohl sind sich alle Akteure über das Ziel einig, möglichst umfassend über germanistische Publikationen zu informieren, und die Nutzer schätzen den Informationsgehalt entsprechend ein.³² Diese Unschärfe vorausgesetzt, haben wir gefragt, welche Autoren mit je eigenen Beiträgen gemeinsam in einer Zeitschrift und zugleich in einer anderen Zeitschrift vertreten sind. Wir vermuten, dass sich daraus erkennen lässt, welche Zeitschrift(en) die Akteure als Publikationsort privilegieren und damit von der Fachcommunity für am wichtigsten gehalten werden. Auf das Netzwerk von Zeitschriften, das sich auf Grundlage von gemeinsamen Autoren ermitteln lässt, wurde das netzwerkanalytische Gruppierungsverfahren der *Community Detection* bzw. *Clique Detection* angewendet.³³ Dieses Verfahren ordnet Zeitschriften einer gemeinsamen Gruppe zu, sofern eine möglichst große Anzahl von Autoren jeweils Beiträge in möglichst

³⁰Eine ausführlichere Auswertung der *Germanistik* durch Jörn Kreutel, Steffen Martus, Erika Thomalla und Daniel Zimmer ist in Vorbereitung. Auch in diesem Fall danken wir dem *DLA* und den o. g. Personen. Die nachfolgend genannten Auswertungen sind im Detail vorbehaltlich einer noch ausstehenden Optimierung der Abgrenzung von Zeitschriften und in Serien erschienenen Sammelbänden. Diese geht aus den Daten der *Germanistik* nicht eindeutig trennscharf hervor, was eine manuelle Klassifikation der Serientitel erforderlich machte. Die hier vorgestellten Analysen und Abbildungen wurden unter Verwendung der Ergebnisse des Netzwerkanalysewerkzeugs *NetworKit* erstellt, das in eine im Aufbau befindliche Analyse- und Visualisierungsumgebung integriert wurde. Vgl. dazu Christian L. Staudt/Aleksejs Sazonovs/Henning Meyerhenke, „NetworKit. A tool suite for large-scale complex network analysis“, in: *Network Science* 4/4 (2016), 508–530; sowie Jörn Kreutel, „Creation and visualisation of network analyses for humanities research. Abstract“, in: *Reader zur ‚Digital Cultural Heritage‘-Konferenz 2017, Berlin, 30.8.–1.9.2017*, 8 f.

³¹Zudem sind die Auswahlkriterien besonders in der Anfangsphase nicht selten kontingent (weil etwa eine Zeitschrift zugänglich ist oder nicht oder weil eine nicht näher bestimmte Auffassung vom Wert eines Periodikums und seiner Beiträge für die bibliographische Aufnahme verantwortlich sind).

³²So zumindest der Eindruck, der sich aus den Ergebnissen einer Umfrage unter Nutzern der *Germanistik* ergibt, die 1968 veranstaltet wurde und deren Dokumente im *DLA*, Marbach a.N., aufbewahrt werden.

³³John Scott, *Social network analysis*, Thousand Oaks/London/New Delhi/Singapur 2017; Staudt/Sazonovs/Meyerhenke (Anm. 31).

vielen Zeitschriften der betreffenden Gruppe veröffentlicht hat.³⁴ Tatsächlich ergab sich aus diesem quantitativen Verfahren eine intuitiv plausible Gruppierung von Zeitschriften, bei der die *DVjs* hinsichtlich verschiedener netzwerkanalytischer Kennzahlen³⁵ auf vorderen Plätzen rangiert und zusammen u. a. mit den folgenden Periodika einer gemeinsamen Gruppe zugeordnet wird: *Zeitschrift für deutsche Philologie*, *Euphorion*, *Germanisch-romanische Monatsschrift*, *Jahrbuch der Deutschen Schillergesellschaft*, *Wirkendes Wort*, *Zeitschrift für deutsches Altertum und deutsche Literatur*, *Beiträge zur Geschichte der deutschen Sprache und Literatur*. Im Gesamtnetzwerk ist diese Gruppe, die wir provisorisch als ‚A-Zeitschriften‘ bezeichnen, wie in Abb. 1 platziert.³⁶

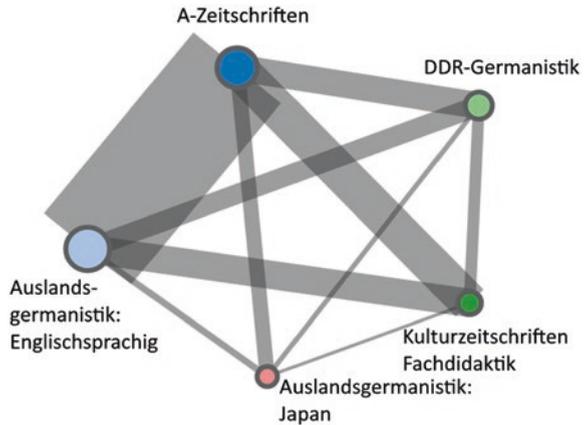
Wir haben uns einer korpusanalytischen Untersuchung der *DVjs* in drei Schritten genähert: In einer ersten Gruppe von ‚Experimenten‘ geht es darum, wie sich die Verarbeitung von Forschung über die Zeit hinweg geändert hat, untersucht auf der Grundlage der Fußnotenmenge. In der zweiten Gruppe werden auf

³⁴ Da es sich bei den genannten Kriterien um graduelle Maße handelt, ist die Gruppenzuordnung mit einer Unschärfe behaftet, die in der Kennzahl der sogenannten ‚Modularität‘ (s. Staudt/Sazonovs/Meyerhenke [Anm. 31]) gemessen werden kann, welche in unserem Fall die Trennschärfe bezeichnet, mit der die zugrunde liegende Menge von Zeitschriften in einzelne enger miteinander verbundene Gruppen untergliedert werden kann. Der gemessene eher niedrige Modularitätswert von 0,29 wird bei näherer Betrachtung nachvollziehbar durch das beobachtbar diversifizierte Publikationsverhalten der Autoren bedingt. Dieses zeigt sich im Fall der Gruppe der Zeitschriften, der die *DVjs* zugeordnet wird, in besonders deutlicher Form. So erscheinen 52 % der Publikationen der Autoren dieser Gruppe in Zeitschriften, die anderen Gruppen angehören, und nur 24 % der Autoren publizieren ausschließlich in Zeitschriften der Gruppe. Entfernt man alle Zeitschriften der Gruppe aus dem Netzwerk, ergibt sich für das verbleibende Zeitschriftennetzwerk eine deutlich höhere Modularität von 0,36. Entfernt man alle Autoren, die in Zeitschriften der Gruppe publizieren, erhöht sich die Modularität noch weiter auf 0,47. Diese Effekte sind für die besagte Gruppe am deutlichsten, d. h. sie trägt als Gruppe am stärksten zur Diffusität des Gesamtnetzwerks bei. Ungeachtet der Unschärfen halten wir die Gruppenzuordnung für fachlich plausibel. So handelt es sich unserer Einschätzung nach z. B. bei der Gruppe um die *DVjs* um eine Gruppe von besonders renommierten Zeitschriften, deren Autoren sich durch ein stark diversifiziertes Publikationsverhalten über die Zeitschriften dieser Gruppe hinaus auszeichnen, also über die Grenzen von Gruppennetzwerken hinaus agieren und damit in hohem Maße das Gesamtnetzwerk prägen.

³⁵ Hinsichtlich der Kennzahl der *Degree-Zentralität* (der Anzahl anderer Zeitschriften, in denen Autoren der *DVjs* publizieren; vgl. dazu Staudt/Sazonovs/Meyerhenke [Anm. 31]) liegt die *DVjs* an zweiter Stelle hinter der *Zeitschrift für deutsche Philologie*, hinsichtlich der Anzahl von Autoren insgesamt an dritter Stelle hinter den *Weimarer Beiträgen* und der *Zeitschrift für deutsche Philologie*. Auch hinsichtlich der stärker das Gesamtnetzwerk berücksichtigenden Kennzahl der *Eigenvektor-Zentralität* (s. ebd.) rangiert die *DVjs* auf dem zweiten Platz, d. h. auch die Zeitschriften, mit denen die *DVjs* verbunden ist, weisen ihrerseits ein hohes Maß an Verbindungen zu den Zeitschriften des Netzwerks auf.

³⁶ Für diese Zeitschriften lässt sich ein 30 m-cors messen, d. h. sie haben mindestens 30 Autoren mit mindestens einer anderen Zeitschrift des Netzwerks gemeinsam; vgl. Scott (Anm. 34). Das abgebildete Netzwerk enthält *alle* Verbindungen zwischen den Zeitschriften, die dem 30 m-core angehören. Insbesondere die dargestellten gruppenübergreifenden Verbindungen liegen zum größten Teil unterhalb der Schwelle von 30 gemeinsamen Autoren.

Abb. 1 Kern des Zeitschriftennetzwerks aufgrund gemeinsamer Autoren 1960–80²⁰



einer sehr einfachen Grundlage, der Länge der Texte und Fußnoten, Schlussfolgerungen über die Veränderungen der Arbeitsweisen gezogen. In der dritten und umfangreichsten Gruppe geht es um die Entwicklung des sprachlichen Verhaltens in der Literaturwissenschaft und um die Frage, ob sich diese Entwicklung als Formierung einer Wissenschaftssprache beschreiben lässt. Uns scheint es wichtig, auch über die gescheiterten Experimente zu berichten, also jene Versuche, die Ergebnisse erbracht haben, die wir nicht sinnvoll interpretieren konnten. Das ist das Thema des vorletzten Abschnitts. Zuletzt versuchen wir, die Erfahrungen in der gemeinsamen Arbeit zu reflektieren, um abstrahiert von unserem direkten Forschungsgegenstand der Frage nachzugehen, wie sich die Arbeit von Literaturwissenschaftlern, die Fachgeschichte betreiben, dadurch verändert, dass sie quantitative Verfahren verwenden.

2 Erstes Experiment: Mengenverhältnisse und die Wahrnehmung von Forschung

In einem ersten Schritt haben wir uns auf die Analyse von Fußnoten konzentriert (Abb. 2). Hintergrund waren zum einen allgemeine Überlegungen zum Status der Anmerkungspraxis: In Anmerkungen artikuliert sich eine Art ‚zweite Stimme‘ wissenschaftlicher Personen, der zum Teil sehr strikte Regeln auferlegt werden (etwa beim Prinzip konsequenter Zitationsweise), die sich aber auch mehr Freiheiten als die ‚erste Stimme‘ des Haupttextes erlauben kann, etwa

²⁰(Liste der Zeitschriftentitel im Anhang). Die Breite der Verbindungslinien ist ein Zeichen für die Menge der Autoren, die in beiden Zeitschriftengruppen publizieren.

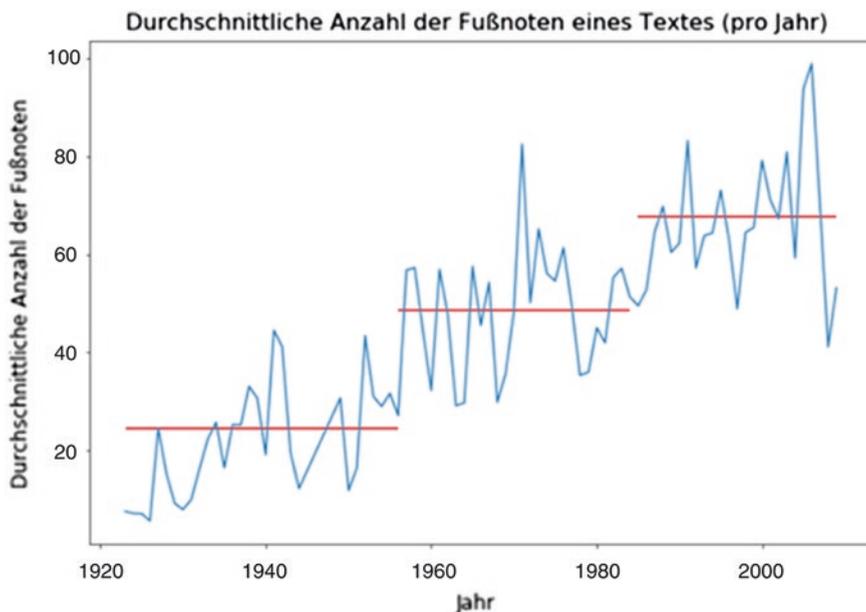


Abb. 2 Durchschnittliche Anzahl der Fußnoten

bei der Formulierung von Mutmaßungen oder persönlicher Kommentare.³⁷ An diesem im Seitenbild ausgewiesenen Ort werden eigene, verwandte und fremde Wissensansprüche modalisiert und vermittelt; hier zeigt sich das Gespür für ein angemessenes und ausreichendes Maß an Dialogizität und Vielstimmigkeit, für die Fähigkeit zur Situierung einer Forschungsfrage im Forschungszusammenhang, zur Perspektivierung ‚epistemischer Dinge‘ und zur Verwaltung jener Grenze, die ‚Wissenschaft‘ von ihren Umwelten trennt und die den Raum markiert, der für Positionierungen einer *Scientific Persona* zur Verfügung steht. Von wesentlicher Bedeutung ist dabei die Dokumentation von Anbahnungsleistungen (Lektüre von Primär- und vor allem Sekundärliteratur), die zur ‚eigenen‘ Erkenntnis geführt haben. Mit einem Wort: Anmerkungen geben wichtige Hinweise auf das Ethos,³⁸

³⁷Zur Räumlichkeit des wissenschaftlichen Schriftbilds und der Platzierung der Fußnoten: Sybille Krämer, „Operationsraum Schrift“. Über einen Perspektivenwechsel in der Betrachtung der Schrift“, in: Gernot Grube/Werner Kogge/Sybille Krämer (Hg.), *Schrift. Kulturtechnik zwischen Auge, Hand und Maschine*, München 2005, 24–57.

³⁸Ralf Klausnitzer/Carlos Spoerhase/Dirk Werle, „Ethos und Pathos der Geisteswissenschaften. Forschungsbericht und Problemskizze“, in: Ralf Klausnitzer/Carlos Spoerhase/Dirk Werle (Hg.), *Ethos und Pathos der Geisteswissenschaften. Konfigurationen der wissenschaftlichen Persona seit 1750*, Berlin/Boston 2015, 13–38, hier insb. 24–31.

mit dem ein Beitrag verfasst wurde.³⁹ Anmerkungen leisten also sehr viel und verdeutlichen *in nuce* die ganze Multinormativität literaturwissenschaftlichen Arbeitens, bei dem es nicht nur um ‚Wahres‘ und ‚Falsches‘ geht, sondern auch um Richtiges, Passendes, Fruchtbare, Interessantes, Spannendes, Relevantes u. v. a. m.⁴⁰ Angesichts dieser Implikationen, so unsere Vermutung, werden selbst scheinbar einfache Daten (Anzahl, Länge, Verteilung etc.) zu implikationsreichen, im Prinzip deutungsbedürftigen Indikatoren.⁴¹

Diese Einschätzung deckt sich mit Befunden des *DVjs*-Archivs: Die Herausgeber nehmen die angemessene Reflexion auf den Forschungsstand sehr wichtig. Immer wieder weisen sie Beiträger darauf hin, dass ihr Aufsatz nicht angenommen wird, weil er unzulänglich mit Sekundärliteraturbelegen ausgestattet ist, oder fordern Autoren von Beiträgen, die sie annehmen möchten, zu Ergänzungen auf. Die Menge der Fußnoten hat also etwas damit zu tun, wie sich ein Beitrag und damit die Zeitschrift insgesamt in der ‚Wissenschaft‘ platziert.

Bei der Analyse der durchschnittlichen Fußnotenzahl eines Textes pro Jahr zeigen sich drei bemerkenswerte Sachverhalte:

1. Die Anzahl der Fußnoten nimmt über den Untersuchungszeitraum zu. Die Entwicklung verläuft nicht kontinuierlich, sondern geschieht in Sprüngen, sodass man von drei Phasen sprechen kann, die in sich relativ homogen sind, während sie sich untereinander signifikant unterscheiden. Homogen bleibt dabei das starke, aber gleichbleibende Schwanken der durchschnittlichen Fußnotenmenge. Der Umfang des Anmerkungsapparats pendelt um einen Mittelwert mit relativ hohen Abweichungen. Zu fragen bleibt, mit welchen Textsorten, Gegenständen, Methoden oder Fachrichtungen (z. B. ÄdL oder NdL) die unterschiedlichen Ausschläge korrelieren. Bei welchen epistemischen Dingen muss eine ‚Forschungslage‘ umfassend mitgeführt werden und wo nicht? Gibt es bestimmte Zugriffe, die als so originell gelten, dass sie aus dem

³⁹Man kann dies gut an den Herausforderungen ablesen, vor denen wissenschaftliche ‚Novizen‘ stehen: Ken Hyland, *Disciplinary Discourses. Social Interactions in Academic Writing*, Ann Arbor 2004, 109–128; Thorsten Pohl, *Studien zur Ontogenese wissenschaftlichen Schreibens*, Tübingen 2007. Zur Fußnotenforschung: Kaspar Brand, ‚Fußnoten und Anmerkungen als charakteristisches Element wissenschaftlicher Darstellungsformen, untersucht am Beispiel der Sprachwissenschaft‘, in: Lutz Danneberg/Jürg Niederhauser (Hg.), *Darstellungsformen der Wissenschaften im Kontrast. Aspekte der Methodik, Theorie und Empirie*, Tübingen 1998, 213–240.

⁴⁰Vgl. dazu Überlegungen zum Wert von ‚Notizen‘ und anderen Formen der Notation von ‚vorläufigen‘ Ergebnissen: Hans-Jörg Rheinberger, ‚Kritzeln und Schnipseln‘, in: Bernhard J. Dotzler/Siegrid Weigel (Hg.), *fülle der combination‘. Literaturforschung und Wissenschaftsgeschichte*, München 2005, 343–356. Im Anschluss daran zur Multinormativität der Literaturwissenschaft: Martus (Anm. 7), 47 f.

⁴¹Martus/Thomalla/Zimmer (Anm. 15).

bewährten Spiel ausscheren und sich den Bezug darauf ersparen dürfen?⁴² Gehen die Abweichungen eher zugunsten oder zuungunsten ausführlicher Anmerkungsapparate, und zeigen sich hierbei im Lauf der Zeit Trends in die eine oder in die andere Richtung?

2. Entscheidend ist für uns zunächst, dass die Fußnotenanzahl innerhalb eines bestimmten Spektrums bleibt. Dies deckt sich mit der Hypothese, dass in einem Fach wie der literaturwissenschaftlichen Germanistik generell sehr viel, aber nicht alles möglich ist, wenn man dazugehören will. Es gibt ein normales Maß an Varianz, das sich jedoch über einen Zeitraum von ca. einem Jahrhundert relativ wenig verschiebt oder sogar annähernd gleichbleibt. Die These radikal zunehmender und daher krisenhafter Diversität bestätigt sich in dieser Hinsicht nicht. Zugleich lässt sich beobachten, dass die Anzahl der Fußnoten generell ansteigt. Offenbar wächst der Erkenntnishaushalt, der mitgeführt, beachtet und verwaltet wird. Man muss mehr im Blick behalten, berücksichtigen und vielleicht auch kognitiv verarbeiten. Bedeutet dies, dass der argumentative Aufwand steigt? Oder entwickeln sich alternative Formen, mit denen etablierte Wissensansprüche verabschiedet und eigene Innovationsansprüche erhoben werden (etwa durch bestimmte Belegstrategien wie kumulative Fußnoten, allgemeine Verweise etc.)? Lassen sich Rückschlüsse auf das Bewusstsein der ‚Fortschrittlichkeit‘ des Fachs ziehen?
3. Zu den bemerkenswertesten Ergebnissen der Fußnotenzählung gehört schließlich der Befund, dass die Quantität nicht kontinuierlich ansteigt, sondern sich in zwei Schüben vollzieht: Ein erster Schub datiert auf die Zeit ‚um 1960‘, ein zweiter Schub ereignet sich ‚um 1990‘. Es liegt nahe, die Hinweise auf wissenschaftshistorische Etappen als Indizien für Modernisierungsprozesse und -phasen zu deuten, die womöglich mit Generationenwechsell zu tun haben. Dies gilt zumindest für die wichtige Transformationsphase der Nachkriegsgermanistik. Diese Entwicklung bezieht etwa Oliver Sill auf umfassende und tiefgreifende „gesellschaftliche[] Modernisierungsprozesse“, in denen die traditionellen Bildungswerte und Bildungstitel, die für den Status der literaturwissenschaftlichen Germanistik in ihrer Phase der Etablierung und Konsolidierung von zentraler Bedeutung waren, an Relevanz einbüßen.⁴³ Diese Veränderungen betreffen diverse Faktoren auf sehr unterschiedlichen Ebenen: die steigende Zahl von Studierenden, der Ausbau des Personals in Forschung und Lehre, der zunehmende Legitimationsdruck angesichts der politischen Kontamination der Fachgeschichte, die Reform der Deutschlehrerausbildung und anderer Studiengänge, die Neubestimmung von Gegenstandsbereichen,

⁴²Lutz Danneberg, „Ich habe nichts Neues zu sagen ...“, in: *Jahrbuch der deutschen Schillergesellschaft* 39 (1995), 434–438.

⁴³Oliver Sill, *Kein Ende und ein Anfang. Germanistische Literaturwissenschaft der sechziger und siebziger Jahre*, Bielefeld 2003, 7 u. 10.

intensivierte methodische Reflexion, Zweifel an der Praxisrelevanz u. v. a. m.⁴⁴ Entsprechend verortet Sill diese historische Gelenkstelle in die Zeit zwischen „1965 und 1980“.⁴⁵

Wir haben an dieser Stelle kein Gegenangebot zu unterbreiten, halten es aber für bemerkenswert, dass sich der von uns festgestellte Veränderungsschub mit Blick auf die Wahrnehmung von Forschung und damit auf die Platzierung im literaturwissenschaftlichen Kommunikationszusammenhang weder zeitlich noch in der beschriebenen Eskalationsdynamik mit dem intuitiv plausiblen Befund von Sill deckt (Befunde, auf die wir unten eingehen, stimmen damit hingegen eher überein). Damit könnte die Stabilität der Phase zwischen ‚um 1960‘ und ‚um 1990‘ darauf hindeuten, dass bestimmte Praktiken, Genres und Medien von institutionellen Veränderungen nicht (direkt) beeinflusst werden und sich zudem gegenüber theoretischen und programmatischen Veränderungen relativ resistent erweisen. So sollten etwa die Evolution des Sozialsystems ‚Wissenschaft‘ und der Institution ‚Universität‘ differenziert behandelt werden.⁴⁶ Dies entspräche der oben referierten Diagnose, der zufolge die Entwicklung der *DVJs* relativ autonom verläuft und sich ‚von außen‘ nicht leicht irritieren lässt. So etwas wie eine ‚Umbruchphase‘ ereignete sich jedenfalls aus der Fußnotenperspektive der Germanistik nicht rund um das magische Datum ‚1968‘. Entsprechend wären weder ältere Bezugskonzepte zu privilegieren, die ‚1945‘ als Zäsur annehmen, noch neuere soziologische und sozialhistorische Modelle, die den entscheidenden strukturgeschichtlichen Wandel auf die 1970er Jahre datieren.⁴⁷ Vielmehr sollten dann eher solche Konzepte priorisiert werden, die in den 1950er Jahren ansetzen.⁴⁸ Wie auch immer man sich entscheidet: Die quantitative Analyse wirft mit Blick auf die Konzeption der germanistischen Wissenschaftsgeschichte in ihren Kontexten zahlreiche Fragen von erheblicher Tragweite auf.

⁴⁴ Ebd., 13.

⁴⁵ Ebd., 7.

⁴⁶ Vgl. dazu und weiteren wichtigen Unterscheidungen: Peter Strohschneider, „Germanistik als Disziplin“, in: Dieter Lamping (Hg.), *Geisteswissenschaften heute. Die Sicht der Fächer*, Stuttgart 2015, 59–73.

⁴⁷ Anselm Doering-Manteuffel unter Mitarbeit von Elisabeth Müller-Luckner (Hg.), *Strukturmerkmale der deutschen Geschichte des 20. Jahrhunderts*, München 2006.

⁴⁸ Beispielsweise Detlef Siegfried, *Time Is on My Side. Konsum und Politik in der westdeutschen Jugendkultur der 60er Jahre*, Göttingen 2006; Elena Agazzi/Erhard Schütz (Hg.), *Handbuch Nachkriegskultur. Literatur, Sachbuch und Film in Deutschland (1945–1962)*, Berlin/Boston 2016. Zur Wissenschaftsgeschichte vgl. Stefan Scherer, „Prägnanz und Evidenz. Philologische Erkenntnis und Verwissenschaftlichung der germanistischen Literaturwissenschaft im disziplinen- und gesellschaftsgeschichtlichen Umbruch der 1950er Jahre“, in: Gerhard Kaiser/Matthias Krell (Hg.), *Zwischen Resonanz und Eigensinn. Studien zur Geschichte der Sprach- und Literaturwissenschaften im 20. Jahrhundert*, Heidelberg 2005, 33–52.

3 Zweites Experiment: Längenverhältnisse und ‚Passung‘ von Beiträgen

Ein letzter Befund zu den Fußnoten betrifft nicht deren Anzahl, sondern deren zunehmende Länge (gemessen wurde die Länge in Anzahl der Zeichen, also Buchstaben, Satzzeichen, Leerzeichen usw.; Abb. 3).

Hier zeigt sich eine klare steigende Tendenz, die zumindest ab 1950 bei ungefähr gleicher Varianz recht kontinuierlich ist und somit keines der bislang erwähnten Diskontinuitätsmodelle stützt. Der Umstand, dass sich hier die beiden Phasen nicht finden, die die Entwicklung der Fußnotenmengen indiziert, wirft die Frage auf, welchen alternativen Entwicklungslogiken die Menge und die Länge von Fußnoten folgen könnten.

Es mag auf den ersten Blick zu trivial erscheinen, sich solche Fragen zu stellen. Die Relevanz von quantitativen Verhältnissen erschließt sich jedoch aus der Perspektive des Archivs. Die dort artikulierten Bewertungspraktiken der *DVjs*-Herausgeber charakterisiert, dass sie bei der Annahme oder Ablehnung ein ganzes Ensemble von Normen zur Geltung brachten, und dies eben auch unter Bezugnahme auf Quantitäten. Die Kriterien gruppieren sich dabei nur lose um Qualitäten wie ‚wahr‘ oder ‚richtig‘. Entscheidend ist das Gespür dafür, welche Beiträge zum Profil der *DVjs* in einem

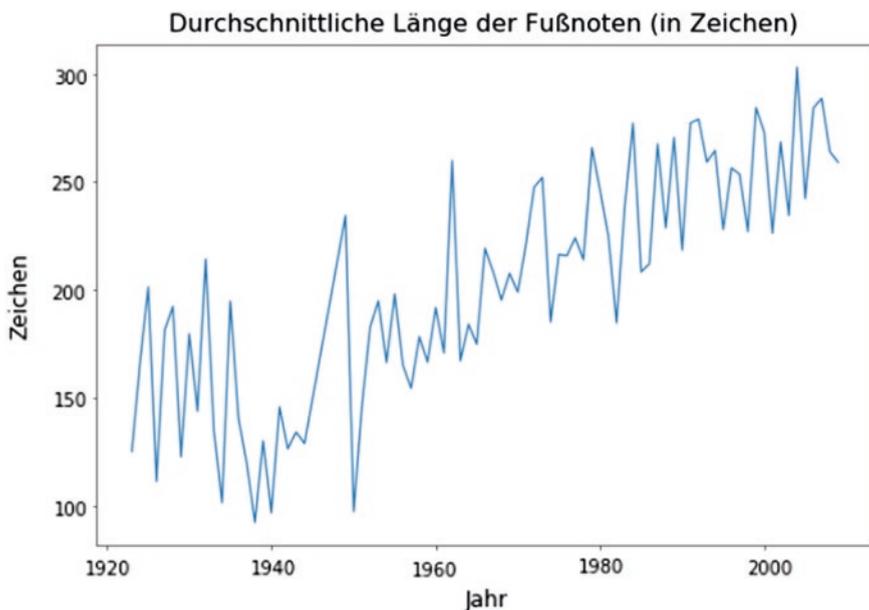


Abb. 3 Durchschnittliche Länge der Fußnoten

Gefüge von Zeitschriften ‚passen‘,⁴⁹ und für diese Einschätzung waren eben auch quantitative Verhältnisse ausschlaggebend. Passungsprobleme⁵⁰ wurden zum Teil dem Gegenstand angelastet: Die Herausgeber stellten dann fest, ein Forschungsobjekt sei zu wenig ‚bedeutend‘, ‚wesentlich‘ oder ‚zentral‘, um angesichts eines Überangebots von Beiträgen in der *DVjs* behandelt zu werden. Zum Teil liegt es aber auch an der Art und Weise, wie mit einem epistemischen Ding umgegangen wurde. Als Hauptkriterium diente dann die Einschätzung der Innovationsleistung eines Aufsatzes, die am bisherigen Forschungsstand gemessen wurde.⁵¹

An dieser Stelle ist die bereits erwähnte Aufmerksamkeit der Herausgeber für Anmerkungen bemerkenswert: Regelmäßig listete Paul Kluckhohn Titel auf, die von einem Beiträger noch verarbeitet werden sollten. Die Entscheidung über die Passung eines Beitrags hing also auch mit (intuitiv bemessenen) quantitativen Kriterien zusammen. Nur bei angemessener Ausstattung mit Fußnoten, so das Argument, ließe sich ein *wissenschaftlicher* Innovationsanspruch erheben. Die Verbindung von ‚Innovation‘ und ‚Wissenschaftlichkeit‘ meint die Einbettung epistemischer Aktivitäten in das Sozial- und Kommunikationssystem ‚Wissenschaft‘. Anmerkungen sorgten dafür, dass die Zeitschrift insgesamt ihren ‚Stellenwert‘ auf dem Feld der literaturwissenschaftlichen Publikationsmedien bewahren konnte. Die Menge und vielleicht auch die Länge von Verweisen war daher für die Annahme oder Ablehnung auch deswegen von Bedeutung, weil es dabei um die Passung für das avisierte Zielpublikum ging: Bemängelt wurde etwa, wenn ein Beitrag sich an ein ‚weiteres Publikum‘ richtete, also an Adressaten, die keine oder zu wenige ‚wissenschaftliche‘ Interessen verfolgten – für diese Gruppe konnte man gern auf Forschungsreferenzen verzichten oder sie nur unvollständig anführen.

Noch in einer zweiten Hinsicht arbeiteten die Herausgeber stetig mit quantitativen Kriterien: Bei den Rückmeldungen an (potenzielle) Beiträger spielte

⁴⁹Dass ein Beitrag abgelehnt wurde, musste also nicht bedeuten, dass er ‚falsch‘ oder wie auch immer ‚schlecht‘ ist, sondern kann auch implizieren, dass der angebotene Aufsatz in einem Spektrum von Alternativen nicht ‚passt‘. Die Relevanz von Passungsverhältnissen zeigt sich auch im Umgang mit Forschern, deren Beiträge ablehnend quittiert werden: Ihnen wird regelmäßig empfohlen, es an einem anderen Publikationsort zu versuchen. Die Abgrenzung kann dabei innerdisziplinär ausgerichtet sein, wenn etwa als Alternative ein Periodikum wie die *Zeitschrift für deutsche Philologie* genannt wird, oder sich auf die Grenze des Sozialsystems ‚Wissenschaft‘ beziehen, wenn populäre Medien oder Publikationsorgane zur Lehrerbildung empfohlen werden. Ein Beitrag zu den Zeitschriftenpraktiken der *DVjs* von Steffen Martus, Erika Thomalla und Daniel Zimmer ist in Vorbereitung. Wir danken dem *DLA*, namentlich Marcel Lepper, Ruth Döring und Simone Waidmann, für die großzügige Unterstützung.

⁵⁰Zur praxeologischen Dimension der Passung vgl. Rahel Jaeggi, *Kritik von Lebensformen*, Berlin 2014, z. B. 108 u. 114.

⁵¹Die Qualifikation von epistemischen Dingen und von Textumgangsformen konnten auch kombiniert werden. In diesem Fall erschien ein Gegenstand nicht hinreichend ‚zentral‘ für seine Behandlung in der *DVjs*, weil sein Ertrag zu wenig Neues bot. Oder ein Beitrag wurde abgelehnt, weil er einen Gegenstand, der eigentlich thematisierungswürdig erschien, nicht ‚erschöpfend‘ genug behandelte.

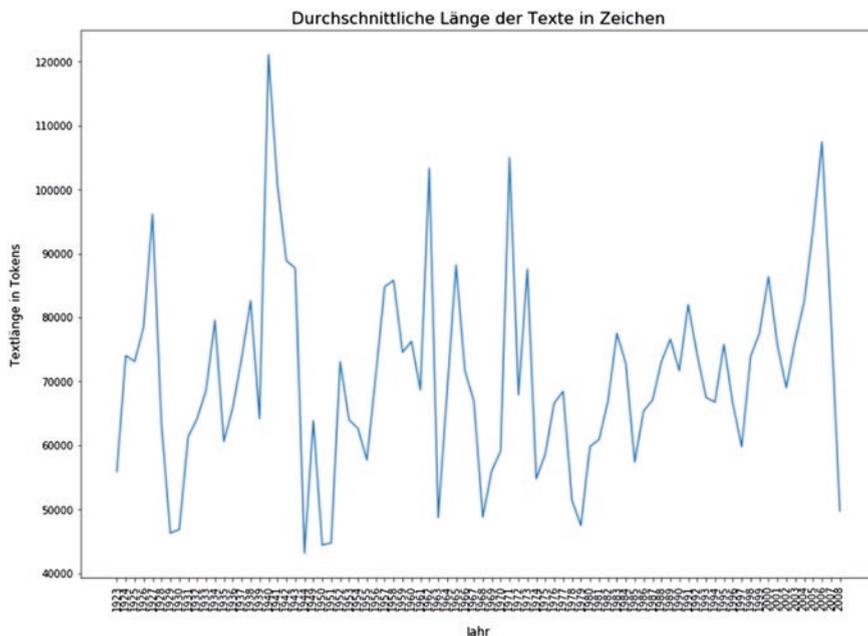


Abb. 4 Durchschnittliche Länge der Texte (in Zeichen)

die Länge der angebotenen Aufsätze eine zentrale Rolle: So gut wie jeder wurde zu Kürzungen aufgefordert. Die Herausgeber reklamierten mithin ein Gespür dafür, wieviel Platz die ‚passende‘ Thematisierung eines ‚passenden‘ Gegenstandes in der *DVjs* in der Regel beanspruchen durfte. Ein Blick auf die durchschnittliche Länge der Texte in einem Jahr zeigt erhebliche Schwankungen (Abb. 4).

Erneut ergibt sich ein bekannter Befund: Die Länge der Aufsätze pendelt trotz großer Schwankungen um ein relativ gleichbleibendes Mittelmaß. Wichtig ist dabei, dass den Herausgebern offenbar Spielräume zur Verfügung standen. Mehr Platz nahmen sie sich z. B. für Themenhefte. Auch Forschungsberichten, die viele der sehr langen Texte (über 300.000 Zeichen) ausmachen, wurde eine Sonderrolle zugestanden.

Aus unseren Untersuchungen ergeben sich mehrere Anschlussfragen. Besonders interessant wäre es, wenn ein systematischer Zusammenhang zwischen der Länge der Texte und anderen Faktoren hergestellt werden könnte. Kandidaten für solche Faktoren sind unserer Einschätzung nach aufgrund der Sichtung der Herausgeber-Korrespondenz und der Untersuchung der ungewöhnlich langen Texte folgende: Zugeständnisse an bestimmte Beiträger (z. B. besonders renommierte oder persönlich eng verbundene Autoren), generische Aspekte (z. B. Sonderstatus der Forschungsberichte), inhaltliche Faktoren (z. B. das Interesse an bestimmten Gegenständen wie etwa hochkanonischen Autoren oder neuen

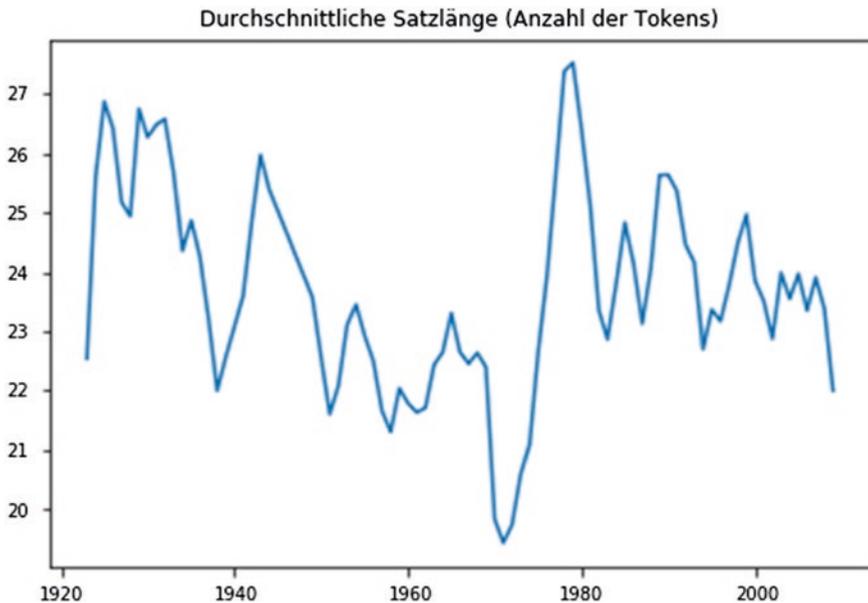


Abb. 5 Durchschnittliche Satzlänge (in Token)

Forschungsobjekten) oder an innovativen Konzepten (z. B. bestimmte Theorien oder Methoden bzw. Zugänge zu Gegenständen, die damit korrelieren).

Eine letzte Beobachtung zu Längenverhältnissen leitet zu unserem dritten Experiment über, das sich mit Indikatoren für die Verwissenschaftlichung germanistischer Fachprosa befasst: Die Satzlänge kann als Hinweis auf die Komplexität der Sprache gesehen werden (Abb. 5).⁵²

Blickt man auf die Entwicklung der Satzlengthen von *DVjs*-Beiträgen, ergibt sich folgendes Bild: Die mittlere Satzlänge scheint in den ersten Jahrzehnten recht stark zuzunehmen, in der Mitte des Jahrhunderts fällt sie dann ab und ab 1970 steigt sie wieder markant an, um von dort an auf einem hohen Niveau zu bleiben, wenn auch mit einer leicht sinkenden Tendenz. Das ist ein erster möglicher Beleg für die grundsätzliche These, die im nächsten Abschnitt diskutiert wird: nämlich dass sich erstens die Wissenschaftssprache der Literaturwissenschaft im Verlauf des 20. Jahrhunderts veränderte und dass sich zweitens in den 1970er Jahren ein Verwissenschaftlichungsschub ereignete, zu dem auch eine komplexere Syntax gehört. Man könnte also aufgrund dieser ersten Datenexploration die These aufstellen, dass sich hier zwei Tendenzen überlagern: Zum einen nimmt die Satzlänge aufgrund der allgemeinen Veränderung der Schriftsprache seit der Mitte des

⁵²Zur Bestimmung der Satzgrenzen und für die Tokenisierung wurde die Python-Bibliothek *NLTK* verwendet: <https://www.nltk.org/> (letzter Aufruf 21.09.2017). Anschließend wurde der Durchschnittswert erst für einen Text und dann für alle Texte eines Jahres ermittelt.

19. Jahrhunderts ab (eine Auswertung von wissenschaftlich-technischen Texten errechnet folgende durchschnittliche Satzlängen: 1850: 32,00 Worte | 1900: 23,58 Worte | 1920: 22,72 Worte | 1940: 19,6 Worte | 1960: 19,9 Worte). Zum anderen entwickelt sich in den 1970er Jahren ein neues Stilideal, das zu ähnlich hohen Satzlängen führt wie am Anfang des 20. Jahrhunderts.⁵³

Auch diese Beobachtung deckt Forschungsbedarf auf: Erstens haben wir hier neben den kontinuierlichen und diskontinuierlichen Progressionsmodellen einen dritten Typus von Veränderung vor uns, sodass sich die Frage stellt, wie diese Entwicklungsformen aufeinander abgestimmt werden können. Zweitens sind die literaturwissenschaftlichen Texte aus den 86 Jahren von 1923 bis 2009 durch mehrere sich überlagernde Tendenzen in der Geschichte der deutschen Sprache bestimmt, und einige der beobachteten Phänomene lassen sich wohl nur unter Berücksichtigung dieses weiteren Kontextes angemessen erklären. Zugleich hat aber – drittens – die Fachsprache der Literaturwissenschaft ihre sehr eigenen Entwicklungsaspekte, zumindest soweit man das aufgrund dieses eingeschränkten Korpus feststellen kann. Darauf wollen wir im Folgenden eingehen.

4 Drittes Experiment: Wissenschaftssprache⁵⁴

Die „Semantik der ‚Szientifizierung‘“ wurde von Rainer Rosenberg in einem einschlägigen Aufsatz behandelt. Im Anschluss an Klaus-Michael Bogdal⁵⁵ diagnostiziert er einen grundlegenden Wandel der Wissenschaftsauffassung und – korrelierend dazu – der Wissenschaftssprache. Im Zuge einer zeittypischen „Wissenschaftseuphorie“ sei die „lebensphilosophische Auffassung von den Geisteswissenschaften“ ebenso verabschiedet worden wie der Typus des „eingeweihten“ Interpreten. Um die „Modernisierung der Disziplin“ zu befördern, habe man den „Anschluß an die internationale Wissenschaftsentwicklung“ gesucht und diesen bei der „Rationalität“, „Methodologie“ und „Begrifflichkeit“ von Sozialgeschichte und Strukturalismus gefunden.⁵⁶ Diese Methodenangebote versprachen „verifizierbare Ergebnisse“ bzw. „empirisch-analytisch lösbare Aufgabenstellungen“ sowie insgesamt eine „Versachlichung“, die sich nicht nur an der

⁵³Die Werte stammen aus einer Studie von Möslein, zitiert nach Karl-Heinz Best, „Kürzungstendenzen im Deutschen aus der Sicht der quantitativen Linguistik“, in: Jochen A. Bär/Thorsten Roelcke/Anja Steinhauer (Hg.), *Sprachliche Kürze*, Berlin/New York 2007, 45–62.

⁵⁴Diesen Abschnitt können wir nicht wie zuvor mit Archivfunden unterfüttern, weil zunächst rechtliche Probleme zu klären sind.

⁵⁵Klaus-Michael Bogdal, „Einleitung. Von der Methode zur Theorie. Zum Stand der Dinge in den Literaturwissenschaften“, in: Ders. (Hg.), *Neue Literaturtheorien. Eine Einführung*, Opladen 1997, 10–31.

⁵⁶Rainer Rosenberg, „Die Semantik der ‚Szientifizierung‘. Die Paradigmen der Sozialgeschichte und des linguistischen Strukturalismus als Modernisierungsangebote an die deutsche Literaturwissenschaft“, in: Ders., *Verhandlungen des Literaturbegriffs. Studien zu Geschichte und Theorie der Literaturwissenschaft*, Berlin 2003, 225–234, hier: 225 f.

abnehmenden Zahl von „Interpretationen“ ablesen lasse, sondern sich vor allem auch in der „Distanz“ zwischen wissenschaftlicher Diktion und „der Sprache des literarischen Textes“ artikuliere.⁵⁷ Während sich noch der Wortschatz der Nachkriegsgermanistik im Wesentlichen aus dem „allgemein[]“ verfügbaren „Bildungsgut“ gespeist habe, favorisierten strukturalistisch und sozialhistorisch gestimmte Wissenschaftler Äußerungen, die auf „Allgemeinverständlichkeit“ ostentativ verzichteten, „sich radikal von der Redeweise der Dichtungsausleger absetzten und [...] auf sprachliche Professionalisierung hinausliefen“ – Erkennungsworte wie „Code, Signifikant, Signifikat, Referent, Denotation, Automatisierung, Verfremdung, literarische Reihe usw.“ kursieren seit dieser Zeit.⁵⁸

Diese Beobachtungen sind ebenso heuristisch fruchtbar wie die Vermutung, dass die „Semantik der ‚Szientifizierung‘“ seit den 1980er Jahren mit der Neigung zum Poststrukturalismus sowie nachfolgend zur Verkulturwissenschaftlichung der Germanistik wieder „relativiert“ werde und „Versuche der Wiederannäherung an den literarischen Diskurs“ zu beobachten seien.⁵⁹ Wie aber lässt sich eine so raumgreifende These operationalisieren? Man wird sie nur prüfen können, wenn man sie in mehrere Aspekte zerlegt. Der linguistischen Forschung haben wir einige Kriterien entnommen, die als typisch für Wissenschaftssprache gelten: Terminologisierung, -ung-Derivate und -bar-Adjektive, dass-Sätze, Vermeidung von ‚ich‘ sowie argumentatives Sprachverhalten.⁶⁰

4.1 Terminologisierung

Es gehört, wie wir oben bereits angemerkt haben, zu den weithin geteilten Einsichten der Wissenschaftsgeschichte der literaturwissenschaftlichen Germanistik, dass um 1960 ein wesentlicher Einschnitt zu verzeichnen ist, der nicht nur den Grad der Verwissenschaftlichung, sondern auch die Terminologie betrifft, dass also „spätestens seit Ende der sechziger Jahre von einem generellen Umbau der Semantik in der bundesdeutschen Literaturwissenschaft die Rede sein kann“.⁶¹

⁵⁷ Ebd., 226 f. Neben neuen Begrifflichkeiten bezieht Rosenberg in seine Überlegungen zur semantischen Modernisierung auch ein, dass sich Begriffsinhalte wandeln oder dass bereits etablierte Begriffe anders gewertet werden (ebd., 229 f.).

⁵⁸ Ebd., 230.

⁵⁹ Ebd., 231–234; Zitate: 232 f.

⁶⁰ Dániel Czicza/Mathilde Hennig, „Zur Pragmatik und Grammatik der Wissenschaftskommunikation. Ein Modellierungsvorschlag“, in: *Fachsprache* 1–2 (2011), 36–59. Die Merkmalsliste der Autoren ist sehr viel länger; sie stellen ausdrücklich fest, dass es kein Modell gibt, um aufgrund der Menge oder der Auswahl der Merkmale festzustellen, dass es sich um einen Fachtext handelt. Uns geht es aber nicht um die Abgrenzung, sondern um die Frage, ob sich die praktische Sprachverwendung der Autoren in der *DVJs* im Laufe der Zeit in Richtung Wissenschaftssprache ändert. Dazu genügt es, wenn wir nur wenige Merkmale überprüfen.

⁶¹ In diesem Punkt stimmt Rosenberg (Anm. 59), 228, mit den zitierten Beiträgen etwa von Bogdal (Anm. 58) oder Sill (Anm. 46) überein.

Auch in diesem Zusammenhang gehen wir von allgemeineren Überlegungen zur Terminologiepraxis aus: Fachbegriffe gelten als Zeichen der Wissenschaftlichkeit.⁶² Sie stehen für Fachkompetenz und bestimmte szientifische Ideale, die sich leicht mit dem Leitcode wahr/falsch verbinden lassen. Es gibt jedoch alternative Einschätzungen, die sich am pädagogischen Leitcode vermittelbar/unvermittelbar orientieren: Nicht selten wird literaturwissenschaftliche Terminologie als unnötiges Vermittlungshindernis behandelt. Im Blick auf Mengenverhältnisse sind daher zunächst die Spielräume bemerkenswert: Bei aller Wertschätzung von Fachbegriffen kann man auch weitgehend ohne literaturwissenschaftliche Terminologie auskommen. Wo aber und wann gilt dies, im Blick auf welche Publikationsformen und Publika?

Zur ersten Erkundung von Terminologiepraktiken haben wir die Lemmaliste der zweiten Auflage des *Reallexikons der deutschen Literaturgeschichte* (1958–1984) und des *Reallexikons für Literaturwissenschaft* (1997–2003) digitalisiert und dann geprüft, wie groß der Anteil der Wörter der Texte eines Jahres ist, der sich dieser Lemmata bedient.⁶³ Unsere These lautete, dass wir einen Rückgang der Begriffe des *RL* 1958 ff. (Anzahl der Lemmata: 401) über den gesamten Zeitraum beobachten können sowie einen Anstieg der Begriffe aus dem *RL* 1997 ff. (Anzahl der Lemmata: 923). Um diesen Effekt zu verdeutlichen, haben wir die Schnittmenge, also die Begriffe, die in beiden Listen zu finden sind, vorher entfernt. Die Ergebnisse bestätigen unsere These nicht (s. Abb. 6).

Die Begriffe des *RL* 1958 ff. nehmen nicht ab; im Gegenteil, es zeigt sich ein fast stetiger Zuwachs der Verwendung nach 1960. Die Begriffe, die nur im *RL* 1997 ff. stehen, zeigen dagegen keine klare Tendenz (s. Abb. 7).

Tatsächlich kann man sich fragen, ob nicht unsere Ausgangshypothese über die Terminologie, die sich in den Lemmata niederschlägt, falsch ist. Zum einen ist unklar, wann die Lemmata der zweiten Auflage des *RL*, deren Bände über ca. 25 Jahre verteilt erschienen, festgelegt worden sind. Vor allem aber dürfte bereits die Lemmaliste des *RL* 1958 ff. als Ausdruck eines Bemühens zu verstehen sein, die Terminologie des Fachs nicht nur retrospektiv zu sichern, sondern auch prospektiv für eine Weiterentwicklung zu sorgen. Mit anderen Worten: Das Lexikon signalisiert den Bedarf an und das Bedürfnis nach fachsprachlicher Veränderung.

⁶²Vgl. zum Folgenden: Michael Kämper-van den Boogaart/Steffen Martus/Carlos Spoerhase, „Entproblematisieren. Überlegungen zur Vermittelbarkeit von Forschungswissen, zur Vermittlung von ‚falschem‘ Wissen und zur Funktion literaturwissenschaftlicher Terminologie“, in: *Zeitschrift für Germanistik* N. F. 21 (2011), 8–24, hier: 19–24.

⁶³Für die Anregung, die Lemmaliste eines literaturwissenschaftlichen Lexikons für unsere Zwecke zu nutzen, danken wir Marcus Willand. Die gezeigten Kurven sind – im Vergleich zu den Originaldaten – in zweierlei Hinsichten bearbeitet: Wir haben Spitzenwerte abgeschnitten, damit die extremen Ausreißer den Trend in den restlichen Daten nicht verdecken. Außerdem haben wir eine Smoothing-Funktion verwendet, die den Trend deutlicher hervortreten lässt (*Savitzky-Golay-Filter*; Fensterlänge: 5, *Polynomial Order*: 2). Um zu vermeiden, dass einzelne Texte, die einen Begriff sehr häufig verwenden, das Gesamtbild verfälschen, wird jeweils nur ausgewertet, ob ein Begriff in einem Text vorkommt oder nicht – jedoch nicht, wie oft er vorkommt.

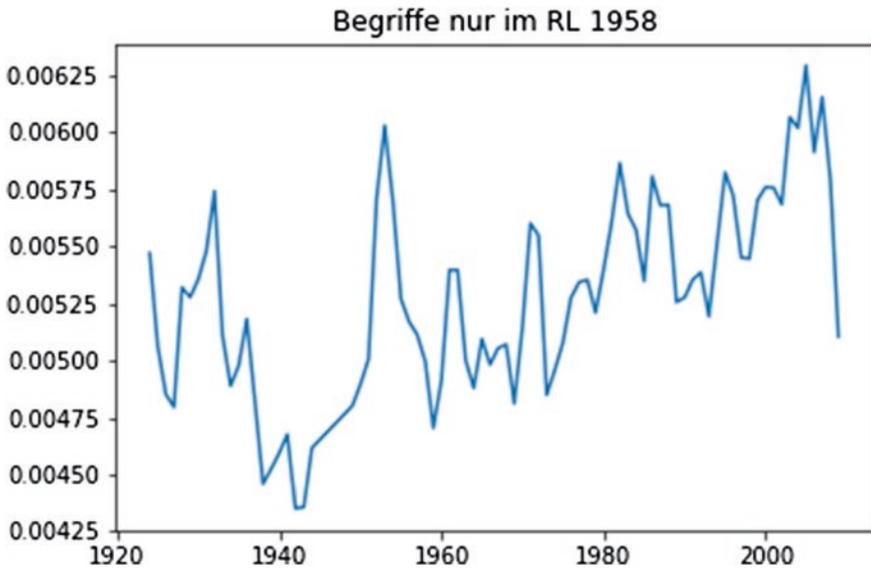


Abb. 6 Begriffe, die nur im RL 1958 ff. vorkommen

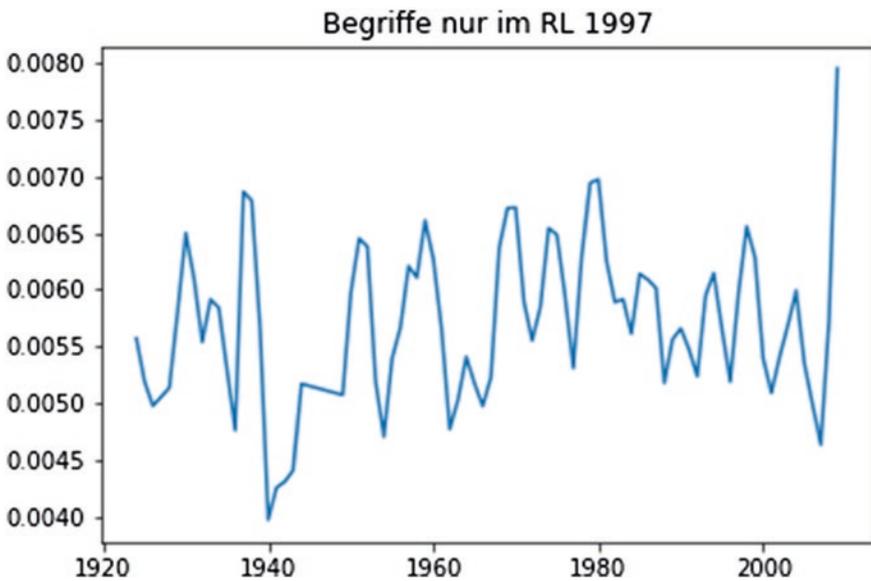


Abb. 7 Begriffe, die nur im RL von 1997 ff. vorkommen

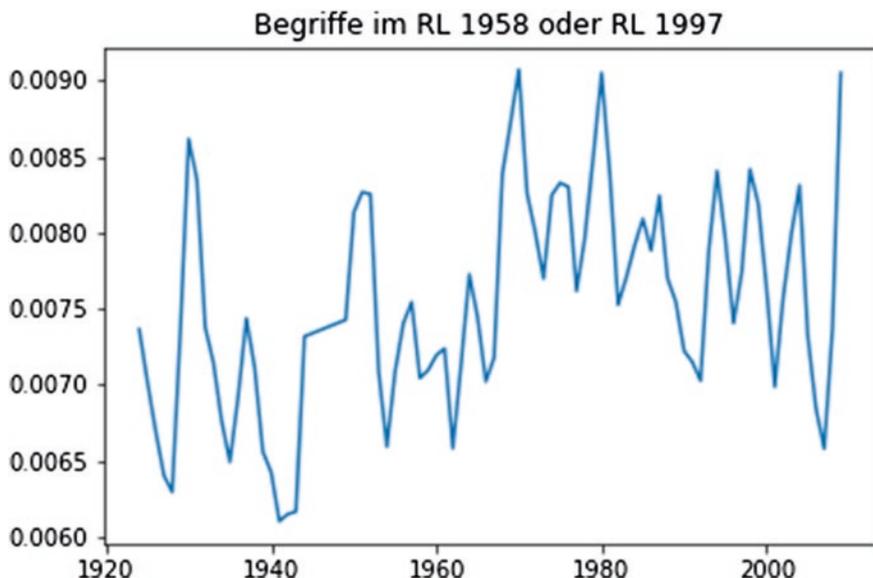


Abb. 8 Begriffe, die in einem der beiden *Reallexika* vorkommen

Daher sollten nicht die Unterschiede, sondern die Gemeinsamkeiten zwischen den Fassungen des *RL* von 1958 ff. und 1997 ff. betont werden. Träfe dies zu, müsste sich eine zunehmende Aufnahme der Terminologie über die Zeit hinweg zeigen, wenn man prüft, wie groß der Anteil der beiden Lemmalisten an den Worten eines Textes ist (Umfang der vereinigten Lemmalisten: 1173).

Was sich in Abb. 8 recht eindeutig ablesen lässt, ist ein signifikanter Anstieg um 1970 und seitdem eine stetige Abnahme auf hohem Grundniveau, während in den Jahren davor nur ein sehr starkes, aber tendenzloses Schwanken der Werte auf niedrigem Grundniveau zu beobachten ist.

Die Schnittmenge der beiden Listen enthält die Begriffe, die ohne Zweifel den Kern der Terminologie des Faches ausmachen – also die Begriffe, die von den 1950er Jahren an bis in die späten 1990er Jahre für zentral erachtet wurden.

Auch hier zeigt sich der Unterschied zwischen der Zeit vor den 1970er Jahren und danach (s. Abb. 9). Man kann also sagen, dass diese Terminologie tatsächlich noch immer wesentlich für das Fach ist, selbst wenn womöglich ein kleiner Rückgang nach 2000 zu erkennen ist.

Eine einfache Gegenprobe gibt einen Hinweis darauf, wo ein Problem unseres Ansatzes liegen könnte. Im Folgenden wird der Anteil der Worte, die oben als typisch für den Strukturalismus genannt wurden („Code“, „Signifikant“, „Signifikat“, „Referent“, „Denotation“, „Automatisierung“, „Verfremdung“), über die Jahre verfolgt (s. Abb. 10).

Anders als die enge Bindung der Begriffe an den Strukturalismus vermuten lassen würde, sehen wir hier, beginnend in den späten 1960er Jahren, tatsächlich

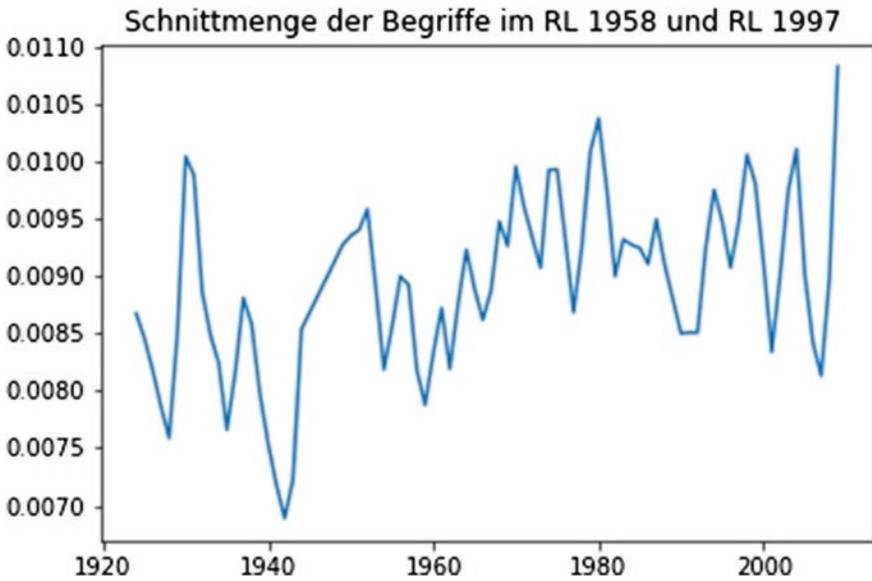


Abb. 9 Begriffe, die im *Reallexikon* 1958 und zugleich im *Reallexikon* 1997 vorkommen

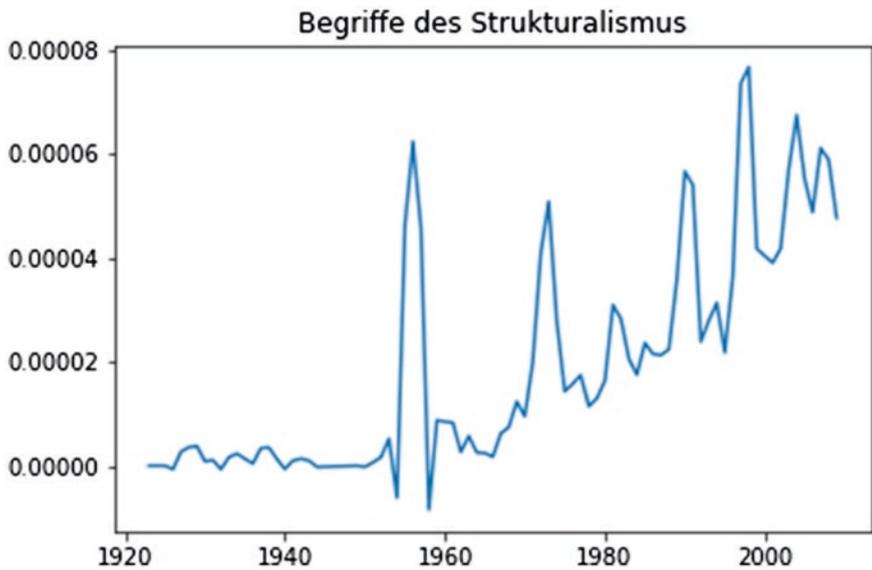


Abb. 10 Typische Begriffe des Strukturalismus. (Nach Rosenberg 2003)

eine ansteigende Verwendung bis in die 2000er Jahre hinein. Das spricht dafür, dass diese Begriffe nicht Teil einer spezifischen Methode sind, sondern in das allgemeine Vokabular des Fachs eingegangen sind.⁶⁴ Allerdings muss man sich, wie ein Blick auf die y-Achse der Grafik zeigt, klar machen: Der Anteil der Begriffe ist relativ klein. Das bestätigt auch ein Blick auf die absoluten Zahlen. Die Begriffe kommen in folgender Häufigkeit vor:

Signifikant: 147 | Referent: 94 | Code: 71 | Signifikat: 16 | Automatisierung: 5 | Denotation: 3 | Verfremdung: 2.

Auch hier wurde zwar, wie oben erläutert, jeder Begriff pro Beitrag nur einmal gezählt, aber seit 1960 sind ca. 1400 Beiträge erschienen. In ähnlicher Weise haben wir uns der poststrukturalistischen Terminologie zugewandt.⁶⁵ Hier steigt die Zahl der Vorkommnisse; die Terminologie wird häufiger und auch nach der Jahrtausendwende genutzt, wenngleich leicht abnehmend (s. Abb. 11).

Wird die Liste der 20 häufigsten Begriffe in den Blick genommen,⁶⁶ dann wird deutlich: Erneut haben einige auch eine allgemeine Bedeutung (z. B. „Philosophie“, „Sinn“, „Schrift“, „Werk“, „Natur“, „Kultur“, „Element“, „Fülle“) bzw. müsste hier die Wortklasse berücksichtigt werden, um sicherzustellen, dass das Wort in seiner terminologischen Verwendung auftaucht (z. B. „selbst“, „macht“).

Spezifische Listen, die charakteristisch für eine Theorie sind, ergeben also sehr viel deutlicher Tendenzen und Entwicklungslinien. Dies gilt auch für das spezifische Untersuchungsvokabular der Narratologie. Im Folgenden zeigt sich, wie die Begriffe der Erzählforschung, ebenfalls ab den 1960er Jahren, zunehmend Bestandteil des literaturwissenschaftlichen Sprachgebrauchs werden,⁶⁷ und zwar angesichts der Werte der y-Achse sehr viel häufiger als die Grundbegrifflichkeit des Strukturalismus (s. Abb. 12).

Auch hier zeigt sich allerdings, dass die häufigsten Worte keineswegs nur terminologisch gebraucht werden:

„geschichte“: 1.609 | „roman“: 1.234 | „welt“: 1.189 | „dauer“: 1.078 | „sprachlich“: 969 | „struktur“: 908 | „geschehen“: 891 | „element“: 838 | „erzählt“: 760 | „systematisch“: 728 | „beziehung“: 688 | „gedanken“: 648 | „gestaltung“: 578 | „situation“: 528 | „realistisch“: 509 | „erzählen“: 503 | „modell“: 408 | „strukturell“: 401 | „verfahren“: 398 | „story“: 395.

⁶⁴Vgl. dazu Rosenberg (Anm. 59), 227.

⁶⁵Grundlage ist eine manuell erstellte Liste von 316 Begriffen aus dem Artikel ‚Poststrukturalismus‘ im *RL* 1997 (Bd. 2, 2003) und der *Wikipedia*.

⁶⁶„selbst“: 3.896 | „sinn“: 3.546 | „spiel“: 1.917 | „macht“: 1.881 | „schrift“: 1.700 | „subjekt“: 1.680 | „werk“: 1.559 | „philosophie“: 1.472 | „ersetzung“: 1.411 | „stimme“: 1.358 | „stimmen“: 1.324 | „raum“: 1.310 | „natur“: 1.249 | „ursprünglich“: 1.239 | „ordnung“: 1.059 | „kultur“: 924 | „fülle“: 905 | „element“: 838 | „unterscheiden“: 788 | „unendlich“: 761.

⁶⁷Grundlage ist eine manuell erstellte Liste erzähltheoretischer Begriffe aus dem Artikel ‚Erzähltheorie‘ im *RL* 1997 und in der *Wikipedia* mit 163 Einträgen.

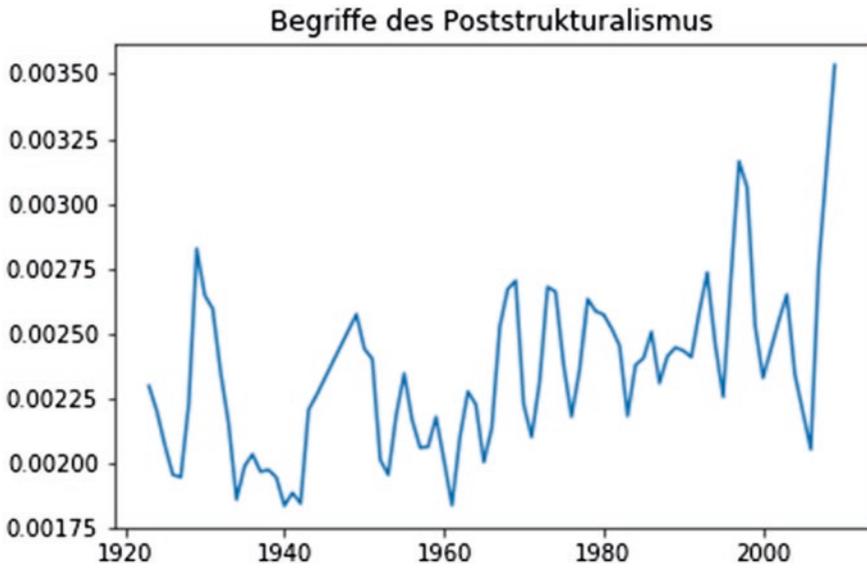


Abb. 11 Typische Begriffe des Poststrukturalismus

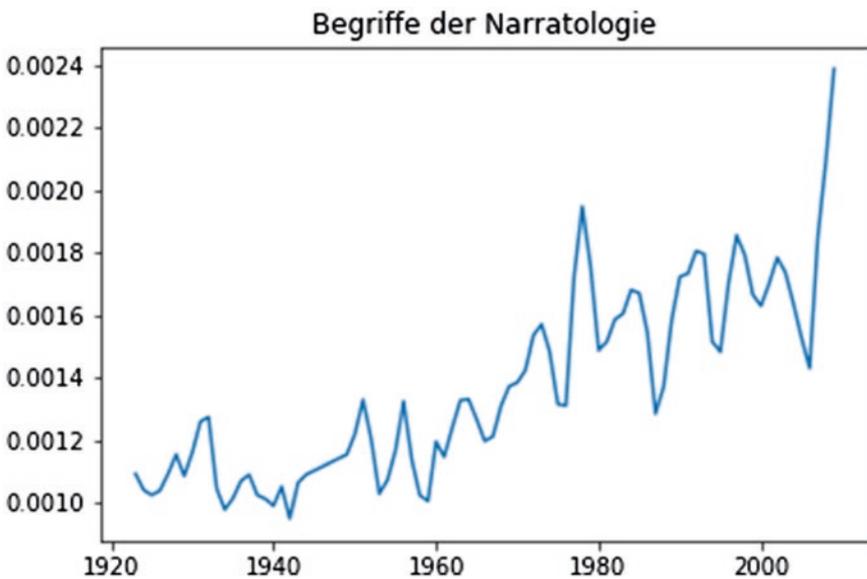


Abb. 12 Typische Begriffe der Erzähltextanalyse

Auf die gesamten Lemmalisten bezogen, ergibt sich daraus die Vermutung, dass deren jeweils kompletter Einsatz wohl zu grob ist: Die Lemmata müssen klassifiziert werden, dabei sollte zumindest etwa zwischen Begriffen der Textanalyse (Metrik, Narratologie usw.), historischen Begriffen (Gattungen, Epochen usw.) sowie Theoriebegriffen unterschieden werden. Ein weiteres Manko des hier verfolgten Ansatzes besteht darin, dass die verwendeten Listen sehr stark von der individuellen Einschätzung eines Philologen abhängen. Um hier allgemeineres Wissen in den Blick zu nehmen, könnte in einem nächsten Schritt von sehr kleinen konsensfähigen Listen mit diskriminativen Begriffen ausgegangen werden, um diese Listen dann mittels *Word-Embedding-Modellen* zur Ermittlung des jeweils nächsten Nachbarn systematisch zu erweitern.

4.2 Weitere Merkmale von Wissenschaftssprache

Neben der Verwendung einer wissenschaftlichen Terminologie gibt es weitere Indikatoren, die für die deutsche Wissenschaftssprache typisch sind. Wie oben erläutert, werden wir hier nur eine kleine Auswahl in den Blick nehmen und dabei feststellen, dass sich ein durchaus widersprüchliches Bild ergibt.⁶⁸ Substantive, die auf ‚-bar‘ enden, bzw. Adjektive, die auf ‚-ung‘ enden, nehmen keineswegs zu; im Fall der Substantive kann wohl sogar von einer Abnahme gesprochen werden (s. Abb. 13).

Anders scheint es sich mit der direkten Vermeidung von Verweisen auf den Autor zu verhalten, also einem der drei ‚Verbote‘ in der Wissenschaftssprache, wie Harald Weinrich es formuliert hat.⁶⁹ Abb. 14 zeigt die Häufigkeit der Wörter „ich“, „mich“, „mein“, „wir“, „unser“, „uns“. Hier kann man die erwartete Abnahme deutlich sehen:

Allerdings ergibt sich sofort ein anderes Bild, wenn die Wortliste aufgespaltert wird und nur die Verweise auf die erste Person Singular in die Zählung einfließen. Hier zeigen sich deutliche Schwankungen ohne Tendenz, d. h. in germanistischen Fachtexten ist das Verweisen auf das ‚Autor-Ich‘ keineswegs weniger üblich geworden (s. Abb. 15).⁷⁰

⁶⁸Die Texte wurden hierfür lemmatisiert; verwendet wurde der *DKpro-Wrapper* mit dem *TreeTagger*. Zum *DKpro-Wrapper* vgl. Fotis Jannidis/Stefan Pernes/Steffen Pielström u. a., „DARIAH-DKPro-Wrapper Output Format (DOF) Specification“, in: *DARIAH-DE Working Papers* 20 (2016). Zum *TreeTagger* siehe Helmut Schmid, „Probabilistic Part-of-Speech Tagging Using Decision Trees“, in: *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK, 1994, 44–49.

⁶⁹Harald Weinrich, „Formen der Wissenschaftssprache“, in: *Jahrbuch der Akademie der Wissenschaften zu Berlin* (1989), 119–158.

⁷⁰Da es beim augenblicklichen Stand der Daten noch nicht möglich ist, die Zitate systematisch aus der Analyse auszuschließen, kann ein Teil der Varianz durch die unterschiedliche Häufigkeit erklärt werden, mit der Texte zitiert werden, die die fraglichen Ich-Verweise enthalten. Allerdings kann man annehmen, dass sich diese Zitate über die Jahre hinweg gleich verteilen, sodass insgesamt ein Trend – sofern er vorhanden wäre – erkennbar sein sollte.

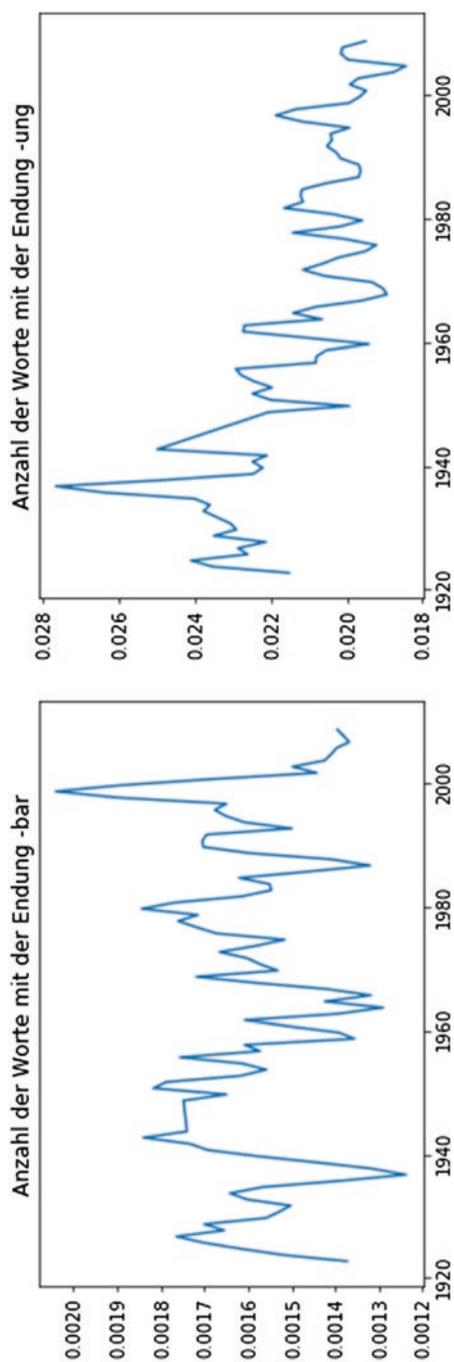


Abb. 13 Entwicklung der Worte mit der Endung ‚-bar‘ bzw. ‚-ung‘

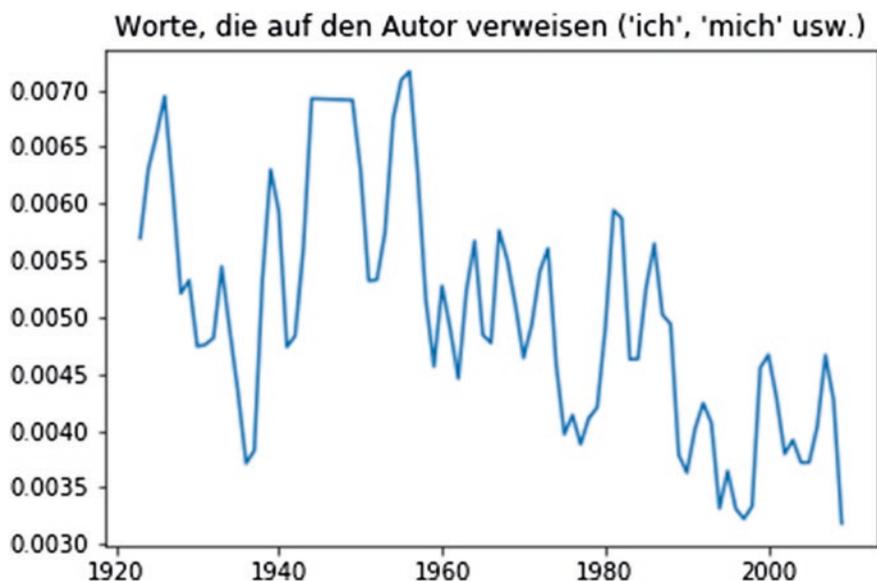


Abb. 14 Entwicklung der Autor-Referenzen („ich“, „mich“, „mein“, „wir“, „uns“, „unser“)

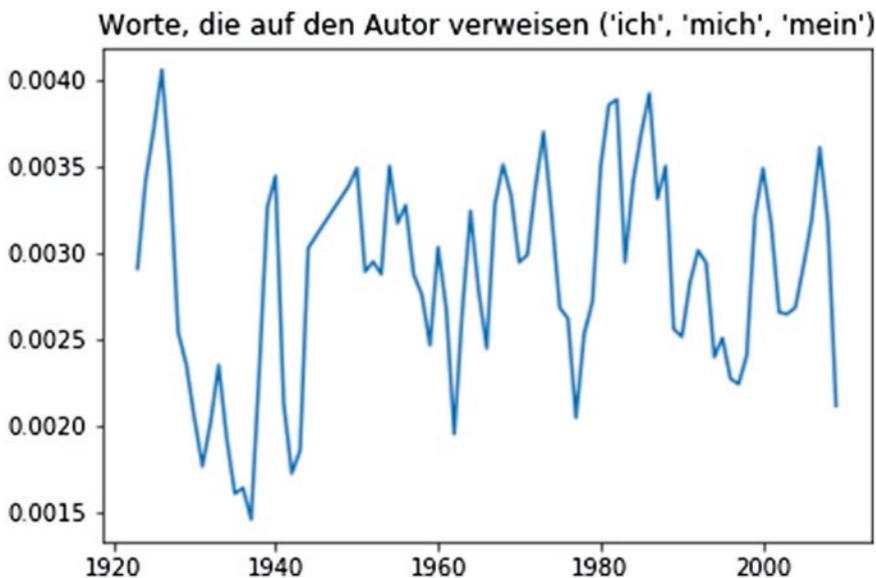


Abb. 15 Entwicklung der Autor-Referenzen („ich“, „mich“, „mein“)



Abb. 16 Entwicklung der Autor-Referenzen („wir“, „uns“, „unser“)

Der Trend, der in der ersten Grafik zu erkennen war, kam durch die Personalpronomina im Plural zustande. Deren Verwendung scheint deutlich aus der Mode gekommen zu sein (s. Abb. 16).

Diese Art der Bezugnahme könnte als altmodisch gelten, sodass ein Geschmackswandel hier womöglich sehr viel wirkungsvoller als das sogenannte deklarative ‚Verbot‘ ist.

Fassen wir zusammen: Von den Merkmalen, die der Wissenschaftssprache zugeschrieben werden und die wir derzeit erfassen können,⁷¹ finden sich in der literaturwissenschaftlichen Sprachverwendung, wie sie sich in den Artikeln der *DVjs* niederschlägt, drei von vier nicht vor: Die Verwendung von ‚ich‘ hat keineswegs abgenommen, die Substantivbildung mit der Endung ‚-ung‘ wird zunehmend

⁷¹ Natürlich gibt es noch zahlreiche weitere Merkmale, die der Wissenschaftssprache allgemein oder der literaturwissenschaftlichen Fachsprache insbesondere zugeschrieben werden, z. B. eine im Vergleich zu anderen Geisteswissenschaften höhere Zahl an „rhetorischen Tropen und Figuren“, s. Andreas Gardt, „Die Fachsprache der Literaturwissenschaft im 20. Jahrhundert“, in: Hugo Steger/Herbert Ernst Wiegand (Hg.), *Handbücher zur Sprach- und Kommunikationswissenschaft*. Bd. 14.1: *Fachsprachen*, Berlin 1997, 1355–1362. Merkmale dieser Art entziehen sich jedoch zurzeit noch einer einfachen Extraktion mit *NLP*-Verfahren. Eine weitere Forschungsperspektive ergibt sich aus der Frage, ob man zusätzliche typische Merkmale, die Wissenschaftssprache aufweist (z. B. Deagentivierung, vermehrte Passiv-Verwendung und eine Zunahme von hypotaktischer Syntax) durch die Verwendung von weiteren *NLP*-Werkzeugen, vor allem einer morphologischen Analyse und einem *Dependency Parser*, nachweisen kann.

gemieden und die Adjektive mit der Endung ‚-bar‘ weisen in ihrer Verteilung keine klare Tendenz auf. Lediglich in der Verwendung von terminologischen Begriffen kann man eine klare Zunahme sehen, auch wenn offensichtlich noch viel Detailforschung notwendig ist, um die Unterschiede zwischen der Verwendung von historischen Begriffen (z. B. der Rhetorik oder ‚Metrik‘) und systematischen Analysebegriffen (z. B. der Erzähltheorie) und Theoriekonzepten (z. B. den Begriffen des Strukturalismus) besser zu verstehen. Um eine so anspruchsvolle These wie die von Rainer Rosenberg zur Szientifizierung der germanistischen Literaturwissenschaft mit Mitteln der Korpusanalyse zu bestätigen, zu modifizieren oder infrage zu stellen, ist es noch zu früh.

5 Experimente ohne Befund

In diesem kurzen Abschnitt soll über die Experimente zumindest berichtet werden, die keine brauchbaren Ergebnisse erbracht haben. Die Bandbreite ist dabei recht groß. Ein wichtiger Bereich betraf den Einsatz stilometrischer Verfahren: So hat die Exploration der deutschsprachigen Daten nicht das erhoffte Clustering nach Teilfächern gezeigt, wahrscheinlich weil es sich hier nicht um stilistische, sondern um rein inhaltliche Unterschiede handelt.⁷² Der Versuch, mit *Burrows' Delta* ein weiteres, ansonsten sehr bewährtes Mittel für die Stilanalyse zu verwenden, führte ebenfalls zu keinem direkt fruchtbaren Ergebnis.⁷³ Ausgehend von der Beobachtung, dass sowohl Theodor W. Adorno als auch Emil Staiger stilistisch wirkmächtige Autoren waren, haben wir ausgewählte Texte der beiden (Adorno: *Noten zur Literatur sowie die Ästhetische Theorie*, Staiger: *Die Kunst der Interpretation*) im Kontext von 180 zufällig ausgewählten Aufsätzen zwischen 1950 und 1985 ausgewertet.⁷⁴ Staigers stilistisch nächster Nachbar war ein Text von Erich Auerbach, sein zweitnächster Nachbar eine Rezension von ihm selbst, danach folgte ein Text von Leif Ludwig Albertsen. Die Texte von Adorno fanden sich nicht in einem Cluster und ihre jeweils nächsten Nachbarn sind in keine für uns sinnvolle Beziehung mit Adorno zu bringen. Die nachfolgende Recherche und das notwendige Feintuning der Daten sind so aufwendig, dass wohl nur in einem eigenen Projekt geklärt werden könnte, ob dieser Ansatz erfolgreich ist.

Ähnlich vorläufig sind unsere Ergebnisse bezüglich der Analyse von argumentativen Indikatoren.⁷⁵ Aus der qualitativen Forschung haben wir eine Reihe von solchen Indikatoren übernommen:

⁷² *Principal Component Analysis (PCA)* der 100 häufigsten Worte mit *scikit-learn*.

⁷³ Vgl. John Burrows: „Delta. A measure of stylistic difference and a guide to likely authorship“, in: *Literary and Linguistic Computing* 17/3 (2002), 267–287.

⁷⁴ Verwendet wurde das R-Script *Stylo* mit den 100 wie auch mit den 3000 most frequent words; vgl. Maciej Eder/Jan Rybicki/Mike Kestemont, „Stylometry with R. A package for computational text analysis“, in: *R Journal* 8/1 (2016), 107–121.

⁷⁵ Hierbei wurden die Wortlisten in Texten von 1960 bis 2009 gesucht.

„weil“ | „daher“ | „also“ | „deshalb“ | „damit“ | „somit“ | „da“ | „denn“ | „wenn“ | „folglich“ | „demzufolge“ | „Folgerung“ | „erschließbar“ | „erschließen“ | „Grund“ | „Gründe“ | „begründen“ | „begründet“.⁷⁶

Allerdings zeigte sich keine Entwicklung der Argumentation über die Zeit. Das änderte sich auch nicht, als diese Liste mittels *Word Embeddings* auf Grundlage der *Wikipedia* von 18 auf 82 Worte erweitert wurde. Das könnte einerseits darauf hindeuten, dass sich die sprachlichen Spuren der Argumentation in diesem Zeitraum nicht änderten, es könnte aber auch sein, dass der Bezug zwischen diesen Worten und der Argumentation als sprachlichem Verhalten zu komplex ist, um durch eine simple Auszählung erfasst zu werden.

6 Einen Schritt zurück

Nachdem wir bislang relativ ungebrochen Wissenschaftsgeschichte betrieben haben, wollen wir zum Abschluss die Frage diskutieren, inwieweit sich unsere Arbeits- und Sichtweise durch *Digitale Literaturwissenschaft* verändert hat. Wir haben uns von unterschiedlichen Seiten dem Thema einer korpusanalytisch verfahrenen Untersuchung der literaturwissenschaftlichen Germanistik genähert. Diese Unterschiede haben unsere Diskussionen bestimmt, aber uns ist nicht ganz klar, wie wir die Effekte gut beschreiben und vielleicht sogar erklären können. Es zeigte sich, dass viele der scheinbar naheliegenden Begriffe eigentlich quer zu oder doch nicht ganz passend für unsere Wahrnehmungen zu sein scheinen. So liegt es etwa nahe, den Unterschied durch die Differenz zwischen ‚hermeneutisch‘⁷⁷ einerseits und ‚empirisch‘⁷⁸ andererseits zu beschreiben. Mit dieser unpassenden Dichotomie verwickeln sich Wissenschaftlerinnen und Wissenschaftler jedoch in unfruchtbare Grabenkämpfe. Denn es ist – allein angesichts der großen Bedeutung von Autopsie – offensichtlich, dass zur philologischen Praxis empirische Teilpraktiken gehören.⁷⁹ Mit dem Abkürzungsbegriff ‚empirisch‘ ist daher in der Regel sehr viel mehr gemeint, als das Wort auf den ersten Blick anzeigt (ein Gefüge von Praktiken, disziplinären Grenzziehungen, Normen u. v. a. m.). Eine weitere Komplexitätssteigerung entsteht dadurch, dass empirisch-quantitative Datenanalyse ihrerseits keineswegs immer nach dem idealtypischen Reglement einer kritisch-rationalen Thesenfalsifizierung vollzogen

⁷⁶Wir danken Simone Winko, die die Argumentation in literaturwissenschaftlichen Texten untersucht, für diese Liste.

⁷⁷Der hier verwendete Begriff des ‚Hermeneutischen‘ ist so weit gefasst, dass er jede auf Lektüre basierende Informationsgewinnung bezeichnet.

⁷⁸Vgl. z. B. die Verwendung dieser Kategorie bei Franco Moretti, „Einleitung. Die vermessene Literatur“, in: Mark Algee-Hewitt u. a. unter der Leitung von Franco Moretti (Hg.), *Literatur im Labor*, Konstanz 2017, 7–16, hier: insb. 8 f., 11 u. 16.

⁷⁹Philip Ajouri/Katja Mellmann/Christoph Rauen (Hg.), *Empirie in der Literaturwissenschaft*, Münster 2013.

werden muss. In der *Distant Reading*-Praxis, wie sie etwa zum Teil in den *Reports* des *Stanford Literary Lab* oder den Arbeiten der *dlina*-Gruppe dokumentiert ist,⁸⁰ zielen Aktivitäten in erster Linie auf eine Exploration der Daten, ohne dabei stets ein empirisches Forschungsdesign zu verwenden, das systematisch auf den Test von Hypothesen setzt.⁸¹

Der Ansatz von Smiljana Antonijević⁸² passt im Vergleich zu einfachen Dichotomien besser und entspricht unseren einleitenden Überlegungen: Sie hat den Einsatz von digitalen Werkzeugen im *Workflow* der Forschung (empirisch) untersucht. Die Arbeit aller *Humanists* lässt sich dieser Feldstudie zufolge schon seit einigen Jahren nicht mehr jenseits des Digitalen denken. Alltäglich ist der Einsatz digitaler Kommunikationsmedien (E-Mail, Skype u. a.); das Schreiben ist weitgehend vom Analogen ins Digitale migriert (nur am Anfang des Forschungsprozesses werden üblicherweise noch handschriftliche Notizen gemacht); die bibliographische Recherche läuft weitgehend über digitale Plattformen; auch bei der Lektüre verzichten nur die Allerwenigsten ganz auf digitale Repräsentationen und Annotationsverfahren. In vielen Fällen geht es dabei um graduelle Unterschiede. Es gibt jedoch einen großen Unterschied zwischen *Humanists* und *Digital Humanists*: den Einsatz von digitalen Verfahren bei der Analyse. Ausgehend von unseren Erfahrungen mit dieser Station des Forschungszusammenhangs versuchen wir im Folgenden, einige Herausforderungen und Spezifika unserer Zusammenarbeit zu beschreiben, wobei wir an dieser Stelle keinen Anspruch auf Systematizität und Vollständigkeit erheben wollen und können.

Ein merklicher Unterschied liegt unseres Erachtens in dem unterschiedlichen Grad an expliziter Prozesshaftigkeit der Forschung.⁸³ In der Philologiegeschichte des 19. und 20. Jahrhunderts sind Mutmaßungen, unabgeschlossene Untersuchungen, offene Fragen, Spekulationen, vorläufige Ergebnisse etc. in der Regel etwas, das im Arbeitszusammenhang der Forschung entweder vor der Publikation seinen Ort hat oder in der Veröffentlichung an ganz bestimmten Stellen (z. B. in Fußnoten, Ausblicken oder auch in mündlichen Äußerungen). Hierbei spielen

⁸⁰ *Stanford Literary Lab*, <https://litlab.stanford.edu/>; *dlina*-Gruppe <https://dlina.github.io/> (letzter Aufruf der Links 21.09.2017).

⁸¹ Etwas passender erscheint uns die Unterscheidung, die mit den Schlagwörtern *corpus-based* und *corpus-driven* verbunden ist. Diese bezieht sich in der Korpuslinguistik herkömmlicherweise auf die Beziehung zwischen Hypothese und Falsifikations- bzw. Validierungsverfahren. Im Fall von *corpus-based* kann die Hypothese ihren Ursprung etwa im Sprachwissen des Forschenden haben oder auf der Grundlage anderer Beobachtungen entstanden sein, während *corpus-driven* den Anspruch erhebt, dass die Hypothese auf der Grundlage des Korpus generiert wird. Beide Verfahren setzen indes voraus, dass eine einmal generierte Hypothese dann systematisch getestet wird, während im Bereich des *Distant Readings* bislang das systematische Testen von Hypothesen nur von einigen Wissenschaftlern propagiert wird.

⁸² Antonijević (Anm. 12).

⁸³ Schruhl (Anm. 16), 47–51.

wohl auch medientechnische Gründe eine Rolle, zumindest gelten die Monographie und der Aufsatz auf Papier immer noch – und über Disziplinengrenzen hinweg⁸⁴ – als die renommierteste Form der Publikation.

Wichtig ist nun, dass an bestimmten Stellen der germanistischen Wissenschaftsgeschichte die explizite Prozesshaftigkeit akzeptiert wurde: Jacob Grimm etwa reklamierte für sich den „Mut des Fehlens“, Wilhelm Scherer schloss daran an und führte den „Anreger“ als Forschertypus ein. Beide wussten, dass sie damit das philologische Ethos provozierten.⁸⁵ Es handelte sich um Situationen, in denen zum einen die Fülle des erschlossenen Materials riskante Forschung probat erscheinen ließ und zum anderen neue analytische Verfahren und entsprechende theoretische Innovationen etablierte „Routinen“ in Frage stellten. Die Akzeptanz vorläufiger Ergebnisse, deren baldige Falsifikation zu erwarten war, basierte dabei auf dem Vertrauen auf Fortschrittlichkeit. Führt mithin die Skepsis gegenüber einer Wissenschaft, die als fortschreitender, vielleicht sogar akkumulierender Prozess konzipiert wird, ebenfalls dazu, dass ein entsprechend hoher Anspruch an die Dauerhaftigkeit des Publizierten gestellt wird?

Prozessualität und damit ein gekonnter Umgang mit Vorläufigkeit und der transitorischen Relevanz der eigenen Forschung ist mithin Teil von moderner Forschung überhaupt. Entscheidend ist indes, wie diese Aspekte in die Forschungspraxis eingehen, d. h. den Arbeitszusammenhang als Set von miteinander verknüpften Teilpraktiken beeinflussen. Das Besondere an den Forschungspraktiken der *Digital Humanities* besteht darin, dass die angeführten Aspekte nicht zu Teilen dominant werden, sondern geradezu kumulieren. Darin könnte ein Grund für die ‚Provokation der Literaturwissenschaft‘ liegen, die viele der allergischen Reaktionen etwa auf Verfahren des *Distant Readings* als aufschlussreiches Symptom erscheinen lassen. In der Praxis einer den *Digital Humanities* nahestehenden oder sich als Teil dieses Forschungszusammenhangs auffassenden Literaturwissenschaft sind nicht-analoge und nicht-monographische Publikationsformen gewöhnlich; sie steht der Herausforderung durch eine überbordende Fülle an Daten gegenüber; sie generiert in Bezug darauf ungewohnte theoretische Perspektiven; sie akzeptiert den produktiven Wert von als solchen markierten ‚Anregungen‘; und sie unterhält ein entsprechend positives Verhältnis zum Fortschreiten der Wissenschaft. Die Vorläufigkeit des jeweiligen Forschungsbeitrags etwa wird von vielen Forschern vertreten, weil auf verschiedenen Ebenen die Übertreffbarkeit und Verbesserbarkeit sehr deutlich erlebt wird: Am Ende einer Forschungssetappe, die für publikationswürdig gehalten wird, kann man zumeist schon deutlich sagen, wo man ansetzen müsste, um das bisher Erreichte eventuell

⁸⁴ Antonijević (Anm. 13), 62.

⁸⁵ Steffen Martus, „jeder Philolog ist eine Sekte für sich“. Wilhelm Scherer als Klassiker des Umgangs mit Klassikern“, in: *Mitteilungen des Deutschen Germanistenverbandes* 53/1 (2006), 8–2; Ders., „Der Mut des Fehlens“, in: Ralf Klausnitzer/Carlos Spoerhase/Dirk Werle (Hg.), *Ethos und Pathos der Geisteswissenschaften. Konfigurationen der wissenschaftlichen Persona seit 1750*, Berlin/Boston 2015, 61–78.

noch zu übertreffen.⁸⁶ Das hat etwas mit der Erstellung und Aufbereitung der Daten zu tun – nicht nur erweitern sich Zug um Zug die zur Verfügung stehenden Korpora, auch deren Erschließung regt zugleich selbst zu immer feineren Beobachtungsleistungen an. Und natürlich bleibt die explosionsartige Fortentwicklung von Algorithmen nicht ohne Wirkung auf das Wissenschaftsethos.

Ein weiterer Punkt, der uns beachtenswert erscheint, betrifft die unterschiedlichen Bedürfnisse, begriffliche und theoretische Klärungen vor der Durchführung der Analyse vorzunehmen, das Ensemble von Praktiken also in einer bestimmten Weise zu arrangieren. Für die Reflexion der eigenen Praxis bei der Verwendung digitaler Verfahren war eine Sequenz von Experimenten für uns besonders aufschlussreich, die die Frage der sprachlichen Szientifizierung betraf: Ausgelöst wurde unser Interesse an der Terminologie u. a. durch die Frage, ob wir auf diese Weise den vermuteten Verwissenschaftlichungsschub Ende der 1960er Jahre und in den 1970er Jahren in einem Wandel der Praxis wahrnehmen können. Wir haben dazu die plausible Idee aufgegriffen, dies über die Verwendung einer Lemmaliste zu operationalisieren. Schon bei der ersten Datenexploration kamen wir jedoch zu dem Schluss, dass der Befund nicht aussagekräftig erschien. Unsere Lösung für dieses Problem, nämlich statt einer Gesamtliste nun Teillisten zu nehmen, lässt sich als neue Operationalisierung begreifen, die einfacher erklärbare Ergebnisse erzeugte.

Warum aber führte der erste Ansatz nicht zum Ziel? Wir haben infolge des fehlgeschlagenen Experiments gesehen, dass wir die Inhomogenität der literaturwissenschaftlichen Terminologie berücksichtigen müssen.⁸⁷ Gerade die unterschiedliche Verwendungshäufigkeit der Begriffe über die Zeit hinweg legt es nahe, den Komplex ‚Terminologie‘ nicht mehr als Einheit zu betrachten. Vorhandene Klassifikationsvorschläge mögen in bestimmten Situationen, z. B. in der Lehre, hilfreich sein, da sie zwar für die Forschungspraxis wenig Relevanz haben, aber doch der Selbstaufklärung sowie der Orientierung des wissenschaftlichen Ethos dienen. Im Zusammenhang mit der quantitativen Forschung verändert sich der Konkretisierungsbedarf jedoch radikal. Die Einsicht in Spezifika von Terminologiegruppen wird höchst relevant, da es sich hierbei um eine Variable handelt, die, wird sie nicht kontrolliert, die Ergebnisse verfälscht und unbrauchbar macht.

⁸⁶Zahlreiche Beispiele bei Schruhl (Anm. 16), 47–51.

⁸⁷Dass die literaturwissenschaftliche Terminologie ganz ausgesprochen inhomogen ist und ihre diversen Teilbereiche u. a. einen sehr unterschiedlichen Grad an Beständigkeit und Klarheit aufweisen, war schon Ausgangspunkt des germanistischen *DFG-Symposiums* 1986, das zu dem weithin wahrgenommenen Band *Zur Terminologie der Literaturwissenschaft* von 1988 führte. Sowohl in dem *Call for Papers* als auch in der späteren Publikation wird die vielschichtige Heterogenität der Terminologie betont: „Ankündigung eines Symposiums zur Terminologie der Literaturwissenschaft“, in: *Zeitschrift für deutsches Altertum und deutsche Literatur* 114/1 (1985), 38–40. Karl Eibl schlägt in seiner Sektionseinleitung dann folgende Klassifikation vor: 1) Technische Begriffe, z. B. der Metrik und Rhetorik, 2) Genealogische Begriffe, vor allem Gattungen, 3) Begriffe für literarische Gruppen, vor allem Epochenbegriffe, sowie 4) Lebensweltliche Grund- und Deutungsbegriffe. Vgl. Karl Eibl, „Einleitung“, in: Christian Wagenknecht (Hg.), *Zur Terminologie der Literaturwissenschaft*, Stuttgart 1988, 357–362.

Dieser Prozess erscheint uns typisch für quantitatives Arbeiten. In den hermeneutischen Alltagsroutinen der Forschung können Begriffe komplex zusammengesetzt sein. Für Eindeutigkeit wird in der Kommunikation aufgrund des Kontextes gesorgt. Im quantitativen Arbeiten aber sind die expliziten Differenzierungen tatsächlich direkt analyserelevant. Das führt schnell dazu, dass diese immer kleinteiliger werdende Forschung sich auf einen Grad der Erkenntnisreichweite beschränkt, der aus der anderen Perspektive nicht unbedingt interessant klingt. Entscheidend ist aber nicht die Frage, was man für mehr oder weniger aufregend hält, sondern dass sich hier ein ganz bestimmtes Gefüge von Teilpraktiken zu einer Praxis formiert: Während Praktiken des Theoretisierens (etwa Klassifizieren, Begriffsklärung etc.) in der breit etablierten literaturwissenschaftlichen Praxis nicht selten ein Eigenleben führen und andere Praktiken (z. B. Analyse) eher vermittelt anweisen oder orientieren,⁸⁸ sind sie im Bereich der quantitativen Analyse unmittelbar folgenreich. An diese Effektivität, verbunden mit Theoriebescheidenheit, muss man sich erst gewöhnen.

Abschließend wollen wir uns, nicht zuletzt aufgrund der hier reflektierten Erfahrungen sowie der erzielten vorläufigen (!) Ergebnisse, in einer im Rahmen der *Digital Humanities* oft diskutierten Frage positionieren: Werden die neuen digitalen Verfahren die analogen ersetzen? Zunächst erscheint uns eine Erinnerung an die großen Szientifizierungsversprechen der 1960er Jahre heilsam. In der Debatte um die Folgen der Digitalisierung für die Literaturwissenschaft kehren die entsprechenden Werte- und Normkonflikte wieder. Damit verbinden sich bestimmte Ideale ‚harter‘ empirischer Forschung (Verifikation bzw. Falsifikation, analytische Grundeinstellung u. a.) sowie ein Habitus der Sachlichkeit, der sich zumindest implizit von ähnlichen Einstellungen und Interessen abgrenzt, wie dies im Rahmen der Szientifizierung der 1960er Jahre bereits der Fall war – hier die ‚alte‘, dort die ‚neue‘ Germanistik. Der Weg führt dann vom einzelnen literarischen Werk zu Texten, die als Teil von größeren Textmengen interessieren; anstelle intensiver Auslegungsarbeit fokussiert sich die Forschung auf Strukturen, Regularitäten und Funktionen.⁸⁹ Das alles ist nicht falsch, aber es hat damals wie heute nie ‚die Germanistik‘ betroffen, weil das Fach noch nie eine homogene Einheit gebildet, sondern unterschiedlichen Arbeitseinheiten Möglichkeiten der Entfaltung geboten hat. Die Zukunft der Literaturwissenschaft wie die einiger anderer geistes- und kulturwissenschaftlicher Fächer dürfte daher eher in einem Neben- und Miteinander von quantitativen und qualitativen Verfahren liegen. In der Arbeit an diesem Text ist jedoch auch deutlich geworden, dass die Praxis der digitalisierten Analyse eine gewisse Eigengesetzlichkeit besitzt. Sie führt dazu, dass es für den individuellen Forschenden nicht immer einfach sein wird, die

⁸⁸ Hier sei an die oben zitierten Studien zur Verwendung von Autor- oder Literaturkonzepten erinnert.

⁸⁹ Bogdal (Anm. 58), 22; Gerhard Lauer, „Die digitale Vermessung der Kultur. Geisteswissenschaften als Digital Humanities“, in: Heinrich Geiselberger/Tobias Moorstedt (Hg.), *Big Data. Das neue Versprechen der Allwissenheit*, Berlin 2013, 99–116.

eine Methode neben der anderen routiniert zu praktizieren. Asymmetrien werden unvermeidlich sein. Damit aber ist ‚die Germanistik‘ schon immer gut zurechtgekommen.

Ergänzung 2021

Auch wenn es inzwischen Plädoyers für quantitative Fachgeschichten gibt,⁹⁰ so sind Beiträge zu einer korpusbasierten Fachgeschichte der Literaturwissenschaften immer noch selten. Das liegt sicherlich auch daran, dass nicht viele Zeitschriften und Monographien vor 2000 maschinenlesbar zur Verfügung stehen. Deshalb ist hier vor allem von den weiteren Arbeiten der Autoren des voranstehenden Beitrags zu berichten.

Eine umfangreiche netzwerkanalytische Studie ist der literaturwissenschaftlichen Bibliographie ‚Germanistik‘ gewidmet.⁹¹ Sie rekonstruiert zudem die Entstehungsgeschichte sowie die Redaktionspraktiken auf der Grundlage bislang unpublizierter Quellen: Die Communities in den Kopublikationsnetzwerken differenzieren sich deutlich entlang von Epochenschwerpunkten (Mediävistik, Barock/Frühe Neuzeit, NDL). Die NDL (ab der Aufklärungszeit) gliedert sich auf der gleichen Betrachtungsebene in mehrere Communities, die eine regionale Differenzierung aufweisen (etwa mit Schwerpunkt im Bereich ‚Österreichischer Literatur‘ oder eine englischsprachige NDL-Community), wie sie für das in unserem Beitrag gezeigte Netzwerk auf Grundlage der Zeitschriftenbeiträge von Autoren erkennbar waren. So existiert auch für dieses Netzwerk eine Community, in deren Zentrum u. a. Zeitschriften stehen, die in Österreich (*Sprachkunst*, *Jahrbuch des Wiener Goethe-Vereins*, *Stifter-Jahrbuch*) und Frankreich (*Études Germaniques*, *Revue d'Allemagne*, *Recherches germaniques*) erscheinen. Die im Beitrag vorgenommene Netzwerkanalyse zeigt, dass eine Betrachtung der von Autoren als Publikationskanäle gewählten Zeitschriften (bzw. der für die Publikation durch Zeitschriftenredaktionen ausgewählten Autoren) ungeachtet der beschriebenen Diversität eine Strukturierung der germanistischen Fachzeitschriften nahelegt. Umgekehrt weisen die Autorencommunities des Sammelbandnetzwerks charakteristische Profile u. a. hinsichtlich der Zeitschriften auf, in denen Beiträge der betreffenden Autoren veröffentlicht werden. Unter der Annahme, dass Sammelbände in der Germanistik mitunter auch ein Tagungsgeschehen und damit – im Gegensatz zu den einzelnen Ausgaben von Zeitschriften – tatsächliche soziale Kontakte abbilden, stellt sich damit für künftige Untersuchungen die Frage, ob und welche Wechselwirkungen zwischen Zeitschriften bzw., Zeitschriftenredaktionen und Tagungen

⁹⁰Arianna Borrelli, „Wissenschaftsgeschichte zwischen Digitalität und Digitalisierung“, in: Martin Huber/Sybille Krämer (Hg.), *Wie Digitalität die Geisteswissenschaften verändert. Neue Forschungsgegenstände und Methoden*, *Zeitschrift für digitale Geisteswissenschaften*, Sonderband 3 (2018), text/html Format. DOI: https://doi.org/10.17175/sb003_001.

⁹¹Jörn Kreutel/Steffen Martus/Erika Thomalla/Daniel Zimmer, „Die Germanistik der Germanistik. Qualitative und quantitative Studien zur Wissenschaftsgeschichte eines „Referatenorgans““, in: *IASL* 44 (2019), 302–379.

bzw. Tagungsorganisatoren hinsichtlich ihrer Rolle als „Gatekeepers“ für den Zugang zu den Scientific Communities der Germanistik bestehen.

Der Beitrag von Jannidis, Konle und Martus setzt die korpusanalytischen Arbeiten im obenstehenden Aufsatz fort.⁹² Den Ausgangspunkt bilden hier fünf Korpora: Deutschsprachige literaturwissenschaftliche Interpretationen, die nach 2000 in der DVjs publiziert wurden sowie Interpretationen aus der DVjs aus den 1960er Jahren; Aufsätze einer sozialwissenschaftlichen Zeitschrift, einer historischen Zeitschrift und einer naturwissenschaftlichen Zeitschrift. Auffällig ist die leichte Separierbarkeit der Texte und zwar nicht nur auf der Grundlage von Inhaltswörtern – was zu erwarten ist –, sondern auch mit Funktionswörtern, ja sogar mit Part-of-Speech-Trigrammen, die eigentlich eher typische syntaktische Regelmäßigkeiten repräsentieren. Die Analyse der Wort- und Satzlänge sowie anderer typischer Stilmerkmale legt einerseits die Vermutung nahe, dass die Literaturwissenschaft sich gleichsam als eine verspätete Wissenschaft in Nachfolge der Sozial- und Naturwissenschaften entwickelt, zeigt aber auf der anderen Seite, dass dieses Bild viel zu einfach ist, vielmehr die Geisteswissenschaften insgesamt wohl eine eigene Entwicklung haben. Blickt man auf die Wortlänge, den Anteil der Substantive oder die Satzlänge, dann finden sich Belege für das Bild des Nachzüglers u. a. im Prozess der Nominalisierung. Andererseits zeigt ein genauerer Blick auf die Substantive, dass die typischen Nominalisierungs-bildungen in der Literaturwissenschaft ab- statt zugenommen haben. Ebenso spricht die – im Vergleich mit den Sozial- und Naturwissenschaften – auffällige Steigerung des Type-Token-Ratio dafür, dass die Entwicklung der literaturwissenschaftlichen Sprache – zumindest auch – von eigenen Tendenzen geprägt ist.

In letzter Zeit wurde vorgeschlagen, Veränderungen der Ähnlichkeit von Teilkorpora in der Wissenschaftsgeschichte als Indikatoren für die Intensität der historischen Umwälzungen zu verwenden.⁹³ Die Grundidee besagt, dass dort, wo die Ähnlichkeit abnimmt, größere Veränderungen vorliegen. Konle, Jannidis, Martus haben versucht, ein geeignetes Verfahren durch Simulationsexperimente zu identifizieren und dieses wiederum auf Texte der DVjs angewendet.⁹⁴ Sie kommen zu dem Ergebnis, dass die Entwicklungsgeschwindigkeit seit den 1950er Jahren zugenommen hat und in den späten 1960er und frühen 1970er Jahren

⁹² Fotis Jannidis/Leonard Konle/Steffen Martus, „Stil der Literaturwissenschaft. Vorläufige Überlegungen aus der Perspektive quantitativer Verfahren (mit einem Seitenblick auf Peter Szondi)“, in: *Zeitschrift für deutsche Philologie*, Sonderheft (2021), 155–182.

⁹³ Stefania Degaetano-Ortlieb/Elke Teich, „Using relative entropy for detection and analysis of periods of diachronic linguistic change“, in: *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, Santa Fe, NM 2018, 22–33.

⁹⁴ Leonard Konle/Fotis Jannidis/Steffen Martus, „Disruptionen der Literaturwissenschaft am Beispiel der DVjs. Methodische Validierung durch Simulation und Anwendung“, in: Manuel Burghardt u. a. (Hg.), *Fabrikation von Erkenntnis. Experimente in den Digital Humanities. Zeitschrift für digitale Geisteswissenschaften*, Sonderband (2021). DOI: <https://doi.org/10.17175/sb005>, 1–31.

die auffälligste Veränderung wahrzunehmen ist, was einerseits fachgeschichtliche Selbstbeschreibungen stützt, andererseits jedoch die Vorstellung einer ständigen Sukzession von grundlegenden Innovationen oder gar von 'Paradigmenwechseln' in Frage stellt.

Anhang

Name	Degree	Eigenvector	Betweenness	
Beitraege_zur_Geschichte_der_deutschen_Sprache_und_Literatur	0.24086	112	0.12276	0.033068
Deutsche_Vierteljahrsschrift_fuer_Literaturwissenschaft_und_Geistesgeschichte	0.49462	230	0.32395	0.075834
Euphorion	0.45376	211	0.28471	0.035927
Germanisch-romanische_Monatschrift	0.41505	193	0.24592	0.032510
Jahrbuch_der_Deutschen_Schillergesellschaft	0.40430	188	0.20513	0.050106
Wirkendes_Wort	0.38495	179	0.21124	0.029847
Zeitschrift_fuer_deutsche_Philologie	0.55269	257	0.35000	0.084607
Zeitschrift_fuer_deutsches_Altertum_und_deutsche_Literatur	0.25376	118	0.15899	0.021615
German_Life__Letters	0.27312	127	0.13552	0.023213
Modern_Austrian_Literature	0.26667	124	0.10562	0.011056
Modern_Language_Notes	0.32258	150	0.15296	0.054417
Modern_Language_Review	0.16559	77	0.074413	0.015572
Monatshefte	0.35054	163	0.20789	0.034326
New_German_Studies	0.094624	44	0.030492	0.0019898
Seminar	0.25376	118	0.12177	0.014233
The_German_Quarterly	0.34194	159	0.22630	0.025354
The_Germanic_Review	0.28387	132	0.17720	0.0096884
Der_Deutschunterricht	0.35269	164	0.12643	0.027743
Merkur	0.24516	114	0.052526	0.025437
Neue_Rundschau	0.29462	137	0.085269	0.041234
Neue_deutsche_Literatur	0.094624	44	0.012387	0.014304
Sinn_und_Form	0.17419	81	0.022097	0.016106
Weimarer_Beitraege	0.27312	127	0.038550	0.067850
Wissenschaftliche_Zeitschrift_der_Friedrich-Schiller-Universitaet_Jena_Thueringen	0.10323	48	0.013126	0.0093494
Doltsu_Bungaku	0.13118	61	0.013358	0.046712
Göte_Nenkan	0.062366	29	0.0050491	0.024260

Literatur

Sämtliche digitalen Referenzen wurden letztmalig am 21.9.2017 eingesehen.

Agazzi, Elena/Schütz, Erhard (Hg.), *Handbuch Nachkriegskultur: Literatur, Sachbuch und Film in Deutschland (1945–1962)*, Berlin/Boston 2016.

Ajouri, Philip/Mellmann, Katja/Rauen, Christoph (Hg.), *Empirie in der Literaturwissenschaft*, Münster 2013.

- Alpaydin, Ethem, *Machine Learning. The new AI*, Cambridge, MA/London 2016.
- Best, Karl-Heinz, „Kürzungstendenzen im Deutschen aus der Sicht der quantitativen Linguistik“, in: Jochen A. Bär/Thorsten Roelcke/Anja Steinhauer (Hg.), *Sprachliche Kürze*, Berlin/New York 2007, 45–62.
- Bogdal, Klaus-Michael, „Einleitung. Von der Methode zur Theorie. Zum Stand der Dinge in den Literaturwissenschaften“, in: Ders. (Hg.), *Neue Literaturtheorien. Eine Einführung*, Opladen 1997, 10–31.
- Borrelli, Ariana, „Wissenschaftsgeschichte zwischen Digitalität und Digitalisierung“, in: Martin Huber/Sybille Krämer (Hg.), *Wie Digitalität die Geisteswissenschaften verändert. Neue Forschungsgegenstände und Methoden*, *Zeitschrift für digitale Geisteswissenschaften*, Sonderband 3 (2018), text/html Format. DOI:https://doi.org/10.17175/sb003_001 https://doi.org/10.17175/sb003_001.
- Brand, Kaspar, „Fußnoten und Anmerkungen als charakteristisches Element wissenschaftlicher Darstellungsformen, untersucht am Beispiel der Sprachwissenschaft“, in: Lutz Danneberg/Jürg Niederhauser (Hg.), *Darstellungsformen der Wissenschaften im Kontrast. Aspekte der Methodik, Theorie und Empirie*, Tübingen 1998, 213–240.
- Burrows, John: „Delta‘. A measure of stylistic difference and a guide to likely authorship“, in: *Literary and Linguistic Computing* 17/3 (2002), 267–287.
- Czicza, Dániel/Hennig, Mathilde, „Zur Pragmatik und Grammatik der Wissenschaftskommunikation. Ein Modellierungsvorschlag“, in: *Fachsprache* 1–2 (2011), 36–59.
- Dainat, Holger: „Anpassungsprobleme einer nationalen Wissenschaft. Die Neuere deutsche Literaturwissenschaft in der NS-Zeit“, in: Petra Boden/Holger Dainat (Hg., unter Mitarbeit von Ursula Menzel), *Atta Troll tanzt noch. Selbstbesichtigungen der literaturwissenschaftlichen Germanistik im 20. Jahrhundert*, Berlin 1997, 103–126.
- Dainat, Holger, „wir müssen ja trotzdem weiter arbeiten‘. Die *Deutsche Vierteljahrschrift für Literaturwissenschaft und Geistesgeschichte* vor und nach 1945“, in: Wilfried Barner/Christoph König (Hg.), *Zeitenwechsel. Germanistische Literaturwissenschaft vor und nach 1945. Eine Veröffentlichung der Arbeitsstelle für die Erforschung der Geschichte der Germanistik im Deutschen Literaturarchiv Marbach am Neckar*, Frankfurt a. M. 1996, 76–100.
- Dainat, Holger/Kolk, Rainer, „Das Forum der Geistesgeschichte“, in: Robert Harsch-Niemeyer (Hg.), *Beiträge zur Methodengeschichte der neueren Philologien. Zum 125jährigen Bestehen des Max Niemeyer Verlags*, Tübingen 1995, 111–134.
- Danneberg, Lutz, „Ich habe nichts Neues zu sagen ...“, in: *Jahrbuch der deutschen Schillergesellschaft* 39 (1995), 434–438.
- Daston, Lorraine, „Die unerschütterliche Praxis“, in: Rainer M. Kiesow/Dieter Simon (Hg.), *Auf der Suche nach der verlorenen Wahrheit. Zum Grundlagenstreit in der Geschichtswissenschaft*, Frankfurt a. M./New York 2000, 13–25.
- Degaetano-Ortlieb, Stefania/Elke Teich, „Using relative entropy for detection and analysis of periods of diachronic linguistic change“, in: *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, Santa Fe, NM 2018, 22–33.
- Div.A., „Historische Praxeologie. Quellen zur Geschichte philologischer Praxisformen. 1800–2000“, in: *Zeitschrift für Germanistik* 23/2 (2013), passim.
- Doering-Manteuffel, Anselm unter Mitarbeit von Müller-Luckner, Elisabeth (Hg.), *Strukturmerkmale der deutschen Geschichte des 20. Jahrhunderts*, München 2006.
- Doerry, Martin, „Schiller war Komponist“, in: *Der Spiegel* 6 (4. Februar 2017), 104–109.
- Drügh, Heinz/Komfort-Hein, Susanne/Koschorke, Albrecht, „Wir Todgeweihten grüßen euch“, in: *Frankfurter Allgemeine Zeitung* 34 (9. Februar 2017), 11, <https://www.faz.net/aktuell/feuilleton/debatten/krise-der-germanistik-antwort-von-heinz-druegh-susanne-komfort-hein-und-albrecht-koschorke-14868192.html>.
- Dubbels, Elke, „Zum Verhältnis von wissenschaftlicher Tradition und Politik im ‚Dritten Reich‘. Die *Deutsche Vierteljahrschrift für Literaturwissenschaft und Geistesgeschichte* in den Jahren 1933–1944“, in: *DVjs* 78 (2004), 672–706.

- Eder, Maciej/Rybicki, Jan/Kestemont, Mike, „Stylometry with R. A package for computational text analysis“, in: *R Journal* 8/1 (2016), 107–121.
- Eibl, Karl, „Einleitung“, in: Christian Wagenknecht (Hg.), *Zur Terminologie der Literaturwissenschaft*, Stuttgart 1988, 357–362.
- Gardt, Andreas, „Die Fachsprache der Literaturwissenschaft im 20. Jahrhundert“, in: Hugo Steger/Herbert Ernst Wiegand (Hg.), *Handbücher zur Sprach- und Kommunikationswissenschaft*, Bd. 14.1: *Fachsprachen*, Berlin 1997, 1355–1362.
- Geulen, Eva, „Für die Einzelsprachlichkeit der Literatur. Nebenbemerkungen zum jüngsten Streit um die Germanistik“, in: *ZfL Blog* (17. Februar 2017), <http://www.zflprojekte.de/zfl-blog/2017/02/17/eva-geulen-fuer-die-einzelsprachlichkeit-der-literatur-nebenbemerkung-zum-juengsten-streit-um-die-germanistik/>.
- Günther, Werner, „Über die absolute Poesie. Zur geistigen Struktur neuerer Dichtung“, in: *DVjs* 23 (1949), 1–32.
- Hyland, Ken, *Disciplinary Discourses. Social Interactions in Academic Writing*, Ann Arbor 2004.
- Jaeggi, Rahel, *Kritik von Lebensformen*, Berlin 2014.
- Jannidis, Fotis/Lauer, Gerhard, „Burrow’s Delta and Its Use in German Literary History“, in: Matt Erlin/Lynne Tatlock (Hg.), *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*, Rochester 2014, 29–54.
- Jannidis, Fotis/Pernes, Stefan/Pielström, Steffen u. a., „DARIAH-DKPro-Wrapper Output Format (DOF) Specification“, in: *DARIAH-DE Working Papers* 20 (2016).
- Jannidis, Fotis/Leonard Konle/Steffen Martus, „Stil der Literaturwissenschaft. Vorläufige Überlegungen aus der Perspektive quantitativer Verfahren (mit einem Seitenblick auf Peter Szondi)“, in: *Zeitschrift für deutsche Philologie*, Sonderheft (2021), 155–182.
- Scott, John, *Social network analysis*, Thousand Oaks/London/New Delhi/Singapur 2017.
- Kämper-van den Boogaart, Michael/Martus, Steffen/Spoerhase, Carlos, „Entproblematizieren. Überlegungen zur Vermittelbarkeit von Forschungswissen, zur Vermittlung von ‚falschem‘ Wissen und zur Funktion literaturwissenschaftlicher Terminologie“, in: *Zeitschrift für Germanistik* N. F. 21 (2011), 8–24.
- Kindt, Tom/Müller, Hans-Harald, „Dilthey gegen Scherer. Geistesgeschichte contra Positivismus. Zur Revision eines wissenschaftshistorischen Stereotyps“, in: *DVjs* 74 (2000), 685–709.
- Klausnitzer, Ralf/Spoerhase, Carlos/Werle, Dirk, „Ethos und Pathos der Geisteswissenschaften. Forschungsbericht und Problemskizze“, in: Ralf Klausnitzer/Carlos Spoerhase/Dirk Werle (Hg.), *Ethos und Pathos der Geisteswissenschaften. Konfigurationen der wissenschaftlichen Persona seit 1750*, Berlin/Boston 2015, 13–38.
- Kluckhohn, Paul/Rothacker, Erich, „Vorwort“, in: *DVjs* 1 (1923), Vf.
- König, Christoph, „Individualität, Autonomie, Originalität. Zur Rezeption Diltheys in den ersten Jahren der *Deutschen Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte*“, in: *DVjs* 67 (1993), 197–220.
- Konle, Leonard/Fotis Jannidis/Steffen Martus, „Disruptionen der Literaturwissenschaft am Beispiel der DVjs. Methodische Validierung durch Simulation und Anwendung“, in: Manuel Burghardt u. a. (Hg.), *Fabrikation von Erkenntnis. Experimente in den Digital Humanities. Zeitschrift für digitale Geisteswissenschaften*, Sonderband (2021). DOI: <https://doi.org/10.17175/sb005>, 1–31.
- Krämer, Sybille, „Operationsraum Schrift“. Über einen Perspektivenwechsel in der Betrachtung der Schrift“, in: Gernot Grube/Werner Kogge/Sybille Krämer (Hg.), *Schrift. Kulturtechnik zwischen Auge, Hand und Maschine*, München 2005, 24–57.
- Kreutel, Jörn, „Creation and visualisation of network analyses for humanities research. Abstract“, in: *Reader zur ‚Digital Cultural Heritage‘-Konferenz 2017, Berlin, 30.8.–1.9.2017*, 8 f.
- Kreutel, Jörn/Steffen Martus/Erika Thomalla/Daniel Zimmer, „Die Germanistik der Germanistik. Qualitative und quantitative Studien zur Wissenschaftsgeschichte eines ‚Referatenorgans‘“, in: *IASL* 44 (2019), 302–379.
- Lauer, Gerhard, „Die digitale Vermessung der Kultur. Geisteswissenschaften als Digital Humanities“, in: Heinrich Geiselberger/Tobias Moorstedt (Hg.), *Big Data. Das neue Versprechen der Allwissenheit*, Berlin 2013, 99–116.

- Martus, Steffen, „jeder Philolog ist eine Sekte für sich“. Wilhelm Scherer als Klassiker des Umgangs mit Klassikern“, in: *Mitteilungen des Deutschen Germanistenverbandes* 53/1 (2006), 8–2.
- Martus, Steffen, „Der Mut des Fehlens“, in: Ralf Klausnitzer/Carlos Spoerhase/Dirk Werle (Hg.), *Ethos und Pathos der Geisteswissenschaften. Konfigurationen der wissenschaftlichen Persona seit 1750*, Berlin/Boston 2015, 61–78.
- Martus, Steffen, „Epistemische Dinge der Literaturwissenschaft?“, in: Andrea Albrecht/Lutz Danneberg/Olav Krämer u. a. (Hg.), *Theorien, Methoden und Praktiken des Interpretierens*, Berlin/Boston 2015, 23–51.
- Martus, Steffen, „Germanistik in der Krise? Der eierlegende Wollmilchgermanist wird dringend gesucht“, in: *Frankfurter Allgemeine Zeitung* 33 (8. Februar 2017), 9, <https://www.faz.net/aktuell/feuilleton/debatten/krise-der-germanistik-antwort-von-heinz-druegh-susanne-komfort-hein-und-albrecht-koschorke-14868192.html>.
- Martus, Steffen, „Wandernde Praktiken ‚after theory‘?“, in: *Internationales Archiv für Sozialgeschichte der deutschen Literatur* 40 (2015), 177–195.
- Martus, Steffen/Spoerhase, Carlos, „Praxeologie der Literaturwissenschaft“, in: *Geschichte der Germanistik* 35/36 (2009), 89–96.
- Martus, Steffen/Spoerhase, Carlos, *Geistesarbeit. Eine Praxeologie der Geisteswissenschaften*, Berlin 2022.
- Martus, Steffen/Thomalla, Erika/Zimmer, Daniel, „Die Normalität der Krise. Beobachtungen zur Geschichte der deutschen Literaturwissenschaft aus Fußnotenperspektive“, in: *DVjs* 89 (2015), 510–520.
- Moretti, Franco, „Einleitung. Die vermessene Literatur“, in: Mark Algee-Hewitt u. a. unter der Leitung von Franco Moretti (Hg.), *Literatur im Labor*, Konstanz 2017, 7–16.
- o. A., „Ankündigung eines Symposiums zur Terminologie der Literaturwissenschaft“, in: *Zeitschrift für deutsches Altertum und deutsche Literatur* 114/1 (1985), 38–40.
- Strohschneider, Peter, „Germanistik als Disziplin“, in: Dieter Lamping (Hg.), *Geisteswissenschaften heute. Die Sicht der Fächer*, Stuttgart 2015, 59–73.
- Pohl, Thorsten, *Studien zur Ontogenese wissenschaftlichen Schreibens*, Tübingen 2007.
- Rheinberger, Hans-Jörg, „Kritzeln und Schnipseln“, in: Bernhard J. Dotzler/Siegrid Weigel (Hg.), *„fülle der combination“. Literaturforschung und Wissenschaftsgeschichte*, München 2005, 343–356.
- Rheinberger, Hans-Jörg, *Experimentalsysteme und epistemische Dinge. Eine Geschichte der Proteinsynthese im Reagenzglas*, Frankfurt a. M. 2006.
- Rosenberg, Rainer, „Die Semantik der ‚Szientifizierung‘. Die Paradigmen der Sozialgeschichte und des linguistischen Strukturalismus als Modernisierungsangebote an die deutsche Literaturwissenschaft“, in: Ders., *Verhandlungen des Literaturbegriffs. Studien zu Geschichte und Theorie der Literaturwissenschaft*, Berlin 2003, 225–234.
- Ryle, Gilbert, *Der Begriff des Geistes*, Stuttgart 1992.
- Scherer, Stefan, „Prägnanz und Evidenz. Philologische Erkenntnis und Verwissenschaftlichung der germanistischen Literaturwissenschaft im disziplinen- und gesellschaftsgeschichtlichen Umbruch der 1950er Jahre“, in: Gerhard Kaiser/Matthias Krell (Hg.), *Zwischen Resonanz und Eigensinn. Studien zur Geschichte der Sprach- und Literaturwissenschaften im 20. Jahrhundert*, Heidelberg 2005, 33–52.
- Schmid, Hans B./Schweikard, David P., „Einleitung. Kollektive Intentionalität. Begriff, Geschichte, Problem“, in: Dies. (Hg.), *Kollektive Intentionalität. Eine Debatte über die Grundlage des Sozialen*, Frankfurt a. M. 2009, 11–65.
- Schmid, Helmut, „Probabilistic Part-of-Speech Tagging Using Decision Trees“, in: *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK, 1994, 44–49.
- Schruhl, Friederike, „Literaturwissenschaftliche Wissensproduktion unter dem Einfluss der Digitalisierung“, in: *Zeitschrift für Germanistik* N. F. 17 (2017), 37–58.

- Schruhl, Friederike: *Formationen der Praxis. Studien zu Darstellungsformen von Literaturwissenschaft und Digital Humanities*, Göttingen 2020.
- Siegfried, Detlef, *Time Is on My Side. Konsum und Politik in der westdeutschen Jugendkultur der 60er Jahre*, Göttingen 2006.
- Sill, Oliver, *Kein Ende und ein Anfang. Germanistische Literaturwissenschaft der sechziger und siebziger Jahre*, Bielefeld 2003.
- Spoerhase, Carlos, „Was ist ein Werk? Über philologische Werkfunktionen“, in: *Scientia Poetica* 11 (2007), 276–344.
- Staudt, Christian L./Sazonovs, Aleksejs/Meyerhenke, Henning, „NetworKit. A tool suite for large-scale complex network analysis“, in: *Network Science* 4/4 (2016), 508–530.
- Tomasello, Michael, *Die kulturelle Entwicklung des menschlichen Denkens. Zur Evolution der Kognition*, Frankfurt a. M. 2006.
- Weinrich, Harald, „Formen der Wissenschaftssprache“, in: *Jahrbuch der Akademie der Wissenschaften zu Berlin* (1989), 119–158.
- Willand, Marcus, „Autorfunktionen in literaturwissenschaftlicher Theorie und interpretativer Praxis. Eine Gegenüberstellung“, in: *Journal of Literary Theory* 5 (2011), 279–302.
- Winko, Simone, „Autor-Funktionen. Zur argumentativen Verwendung von Autorkonzepten in der gegenwärtigen literaturwissenschaftlichen Interpretationspraxis“, in: Heinrich von Detering (Hg.), *Autorschaft. Positionen und Revisionen*, Stuttgart/Weimar 2002, 334–354.
- Winko, Simone/Jannidis, Fotis/Lauer, Gerhard, „Geschichte und Emphase. Zur Theorie und Praxis des erweiterten Literaturbegriffs“, in: Jörn Gottschalk/Tilmann Köppe (Hg.), *Was ist Literatur? Basistexte Literaturtheorie*, Paderborn 2006, 123–154.

Online-Ressourcen

dlina-Gruppe <https://dlina.github.io/>.
 GitHub, Fotis Jannidis, <https://github.com/fotisj>.
 langdetect, <https://pypi.python.org/pypi/langdetect>.
 NLTK, <https://www.nltk.org/>.
 scikit-learn, <https://scikit-learn.org>.
 Stanford Literary Lab, <https://litlab.stanford.edu/>.
 Text Encoding Initiative, <http://www.tei-c.org>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Lieder als kollektive Texte. Digitale Verschriftung und populäre Philologie auf *Genius*

Alexander Nebrig

Neue Medien erweitern das Spektrum der Reproduktionsmittel von Inhalten, verdrängen dabei alte Medien oder konkurrieren mit ihnen. Ein neues Medium kann Inhalte und kulturelle Praktiken ans Licht bringen, die zuvor nur in Räumen einer restringierten Öffentlichkeit zu beobachten waren: etwa ein Konzert, das durch den Rundfunk übertragen wird. Neben einem reproduktiven Mehrwert kann es zudem generativ sein, d. h. die Möglichkeit, neue Inhalte und Formen hervorzu- bringen, kann folglich vom neuen Medium selbst induziert sein.

Auf Sprach- und Textmedien bezogen, meint dies, dass nicht nur bestehende Korpora umformatiert, sondern auch gänzlich neue generiert werden. Die Erfindung der Schrift hat neben die orale Poesiekultur ein skripturales Gegenstück gestellt. Große Teile der mündlichen Kultur wurden seit der Antike aufgeschrieben und auf chirographischen Buchrollen und Kodizes sowie in gedruckten Büchern bis in die Gegenwart als Literatur tradiert. In der Literaturgeschichte bildeten die kontinuierlichen Format- und Medienwechsel genuin literarische Gattungen wie den Roman aus, brachten aber auch immer sprachliche Praktiken bestimmter sozialer Schichten wie das Volkslied ans Licht der Literatur, die bis dato im Schatten einer performativen Mündlichkeit existiert hatten.

Im gegenwärtigen Wandel von analogen zu digitalen Medien tritt die performative Populärkultur des Rap in einen Wahrnehmungskreis, der deutlich größer als jener der Fankultur ist und epistemische Relevanz besitzt. Das Liedgut der populären Formation hat nicht einfach nur wissenschaftliche Aufmerksamkeit

A. Nebrig (✉)
Heinrich-Heine-Universität, Düsseldorf, Deutschland
E-Mail: nebrig@hhu.de

erhalten,¹ weil es zu lange vernachlässigt worden wäre, sondern weil es schlichtweg erst seit Kurzem in interaktiven und nutzergenerierten Hör- und Sehmedien wie *YouTube* präsent ist. Die Mediatisierung ist zum einen audiovisuell; zum anderen aber auch, da es sich primär um ein sprachliches Phänomen handelt, skriptural. Der digitale Medienbruch hat die Liedkultur des Rap als Schrift lesbar werden lassen. Die Überführung ins Medium der Schrift, d. h. die Literarisierung, erfolgte dabei nicht von außerhalb durch wenige Akteure einer professionellen Kritik mit einem Faible für Populärkultur, sondern durch das Publikum selbst, an das sich die populärkulturellen Formen richten.

Zwar wurde die Rap-Kultur von ihren Anfängen an ediert,² aber erst seit den 2000er Jahren laden soziale Medien die gesamte Hörerschaft zur editorischen und auch kritischen Kommunikation ein.³ Die transmediale und interaktive Struktur des Internet hat ein literarisches Korpus hervorgebracht, das in den analogen und gedruckten Medien des populärkulturellen Liedes keine quantitative Entsprechung findet.

War im Laufe des 19. Jahrhunderts der kleine Kreis der Leserschaft auf die gesamte Bevölkerung ausgedehnt worden, unterbreitet der gegenwärtige digitale Medienbruch dieser Gesamtleserschaft das Angebot zur kritischen Teilhabe.⁴ Wolfgang R. Langenbacher, der die zweite, quantitative Leserevolution ins letzte Drittel des 19. Jahrhunderts verlegt, interpretiert sie als Demokratisierung des Lesens.⁵ Passive Teilhabe am gedruckten Wort würde den Befund nüchterner bezeichnen, da mediale und politische Teilhabe nicht notwendig konvergieren. In

¹Markus Heide, *Rap zwischen Mündlichkeit, Schriftlichkeit und Technologie. Eine Verortung anhand deutschsprachiger Raplieder*, Erlangen 2011. Heide hat seine Magisterarbeit bei *Genius* hochgeladen, sodass die kritische Arbeit wie ein Lied annotiert werden kann (<https://genius.com/albums/Markus-heide/Rap-zwischen-mundlichkeit-schriftlichkeit-und-technologie>, letzter Aufruf 24.7.2018); Fabian Wolbring, *Die Poetik des deutschsprachigen Rap*, Göttingen 2015; Johannes Gruber, *Performative Lyrik und lyrische Performance. Profilbildung im deutschen Rap*, Bielefeld 2016; John Gray, *Hip-Hop Studies. An international bibliography and reference guide*, New York 2016.

²Stanley, Lawrence A. (Hg.), *Rap. The lyrics*, New York 1992.

³Zu diesem Konzept und seinen verschiedenen Aspekten ausführlich Steffen Martus, *Werkpolitik. Zur Literaturgeschichte kritischer Kommunikation vom 17. bis ins 20. Jahrhundert mit Studien zu Klopstock, Tieck, Goethe und George*, Berlin/New York 2007, 52–112.

⁴Von kritischer Kollektivität spricht Carolin Wiedemann, *Kritische Kollektivität im Netz. Anonymous, Facebook und die Kraft der Affizierung in der Kontrollgesellschaft*, Bielefeld 2016. Als poetologisches Konzept freilich ist die kollektive Autorschaft in der Romantik vorhanden, vgl. Michael Gamper, „Kollektive Autorschaft/Kollektive Intelligenz. 1800–2000“, in: *Jahrbuch der deutschen Schillergesellschaft* 45 (2001), 380–403.

⁵Wolfgang R. Langenbacher, „Die Demokratisierung des Lesens in der zweiten Leserevolution“, in: Herbert Göpfert/Ruth Meyer/Ludwig Muth u. a. (Hg.), *Lesen und Leben*, Frankfurt a.M. 1975, 12–35.

diesem Sinne wird die digitale Medienkultur als Partizipationskultur etikettiert.⁶ Wie schon für die Lese- und Lesegeschichte gilt auch hier, dass mediale Teilhabe nicht notwendig Demokratisierung bedeutet.

Die alle Lebensbereiche erfassende Laienkritik hat längst nicht nur die Literaturkritik des Feuilletons,⁷ sondern auch die philologische Kritik der Fachwissenschaften erreicht. Deren Teilnehmer sind mit der Frage konfrontiert, ob und wie sie die Amateure einbinden. Vor allem die durch das textkritische Verhalten der Laien entstandenen Korpora stellen eine Herausforderung dar. Denn Laienphilologen laborieren nicht nur an der analogen Schriftüberlieferung, sondern bilden neue Korpora, die aus den medialen Gesetzmäßigkeiten des digitalen Raumes hervorgehen. Ihre kollaborative Praxis und ihre transmediale Perspektive bringen nicht nur neuartige, in bisherigen Kategorien schwer zu fassende Formen des ästhetischen Urteils hervor,⁸ sondern auch formal wie inhaltlich neuartige ‚Literatur‘. Durch kollektive Verschriftlichung und kritische Erschließung des populärkulturellen Liedgutes entstand auf der Wissensdatenbank und interaktiven Plattform *Genius* (genius.com) ein derartiger epistemischer Gegenstand.⁹

Am Anfang stand die Idee, anglophone Rap-Songs, die nicht nur bereits Muttersprachlern schwer verständlich sind, sondern auch als orales Phänomen kaum transkribiert worden waren, verfügbar und für jedermann kommentierbar zu machen. Rap-Liebhaber (Amateure) versorgten nicht mehr nur unilateral andere Freunde dieser Lieder mit Texten. Ebenso regte der Charakter des interaktiven Mediums die Nutzer an, an der Einspeisung von Texten und ihrer Kommentierung teilzunehmen. Seit 2009 ist eine Sammlung von Texten entstanden, die bereits Ende 2018 von

⁶ Henry Jenkins, *Fans, Bloggers, and Gamers. Exploring Participatory Culture*, New York 2006.

⁷ Vgl. Thomas Wegmann, „Warentest und Selbstmanagement. Literaturkritik im Web 2.0 als Teil nachbürgerlicher Wissens- und Beurteilungskulturen“, in: Matthias Beilein/Claudia Stockinger/Simone Winko (Hg.), *Kanon, Wertung und Vermittlung. Literatur in der Wissensgesellschaft*, Berlin 2012, 279–291; Thomas Ernst, „User Generated Content‘ und der Leser-Autor als ‚Prosumer‘. Potenziale und Probleme der Literaturkritik in Sozialen Medien“, in: Christina Gansel/Heinrich Kaulen (Hg.), *Literaturkritik heute. Tendenzen – Traditionen – Vermittlung*, Göttingen 2015, 95–111.

Die Entwicklung der Laienkritik wird kontinuierlich durch Sammelbände begleitet. Hier nur die jüngeren: Andrea Bartl/Markus Behmer (Hg.), *Die Rezension. Aktuelle Tendenzen der Literaturkritik*, Würzburg 2016 und Sebastian Böck (Hg.), *Lesen X.0. Rezeptionsprozesse in der digitalen Gegenwart*, Göttingen 2017.

⁸ Sianne Ngai, *Our Aesthetic Categories Zany, Cute, Interesting*, Cambridge, MA, 2012.

⁹ Ben Fields/Christophe Rhodes, „Listen To Me – Don’t Listen To Me. What Communities of Critics Tell Us About Music. ISMIR 2016“, in: Michael I. Mandel/Johanna Devaney/Douglas Turnbull/George Tzanetakis (Hg.), *Proceedings of the 17th International Society for Music Information Retrieval (ISMIR) Conference 2016, New York City, August 7–11, 2016*, New York 2016, 199–205, https://wp.nyu.edu/ismir2016/wp-content/uploads/sites/2294/2016/07/173_Paper.pdf (letzter Aufruf 24.7.2018).

über 100 Mio. Besuchern monatlich genutzt wurde.¹⁰ Ende 2014 waren es noch 700.000 angemeldete Nutzer gewesen, die über 120.000 Songs eingespeist hatten mit drei Millionen Annotationen; nur zehn Prozent der Nutzer sind aktiv.¹¹ Schon früh hatte sich gezeigt, dass in einer nutzergenerierten Wissensdatenbank subjektive Vorlieben auf Dauer nicht den Maßstab bilden können. Denn nicht mehr nur Rap-Texte, sondern die Texte von Pop-Songs, ja von Volksliedern und Gedichten, Reden, Dramen und Romanen gingen in die Sammlung ein. Auch die Beschränkung auf die anglophone Kultur wurde aufgegeben. Die meisten Nutzer kommen aus dem nord-amerikanischen Raum, es gibt daneben weitere Gemeinschaften wie die deutschsprachige, die frankophone oder die polnische. Die schriftliche Kommunikation über mündliche Poesie ist also tendenziell global angelegt. Darüber hinaus besteht ein zweites Ziel darin, einen Annotationsraum für das gesamte Web zu etablieren.¹² Auch die Selbstbegründung des Vorhabens durch einen Vergleich mit der *Talmud-Exegese*¹³ verweist auf höhere Ziele. Aus dem *Rap-Genius*, so der ursprüngliche Name, wurde „the world’s biggest collection of song lyrics and crowdsourced musical knowledge“.¹⁴

Seit der Investition des Risikokapitalgebers Ben Horowitz ist die auf eine studentische Initiative zurückgehende Wissensdatenbank ein Wirtschaftsunternehmen, das durch Anzeigen finanziert wird und seit 2016 beispielsweise seine Daten an den Musikstreaming-Dienst *Spotify* verkauft.¹⁵ Das wirtschaftliche Interesse am populärkulturellen Wissen beweist, dass der Prozess der Verschriftung nicht sekundäres Beobachterphänomen ist. Die laienphilologische Beobachtung in der Rap-Kultur gehört zu jenen ubiquitären Vorgängen, in denen Rezipienten – ob nun Bürger oder Spieler, Nutzer oder Konsumenten, Hörer, Leser oder Zuschauer – den kreativen, produktiven und distributiven Vorgang systematisch mitgestalten. Rezeption produziert Daten und regt Aktivität zugleich

¹⁰ Andrea Lo, „How a rap lyric inspired a multimillion-dollar online platform“, in: *CNN Money* (29.6.2018), <http://money.cnn.com/2018/06/27/smallbusiness/genius-annotation-platform/index.html> (letzter Aufruf 24.7.2018).

¹¹ Fields/Rhodes (Anm. 9), 199 u. 204.

¹² Vgl. <https://genius.com/web-annotator> (letzter Aufruf 24.7.2018). Über den Browser kann man auf den besuchten Seiten als angemeldeter Nutzer Kommentare für andere angemeldete Nutzer hinterlassen. Für weitere Informationen siehe Reeves Wiedeman, „Genius Idea“, in: *New York Magazine* (29. Dezember 2014, online 4. Januar 2015), <http://nymag.com/daily/intelligencer/2014/12/genius-minus-the-rap.html> (letzter Aufruf 24.7.2018).

¹³ „Horowitz compares the potential of Rap Genius, which already hosts some political speeches, law cases, and *The Great Gatsby* to a ‚Talmud for the Internet.‘ Just as the Talmud is the source of many religious beliefs that aren’t necessarily mentioned directly in the bible, it could one day be the source of ‚knowledge on the knowledge‘ online.“ Vgl. Sarah Kessler, „Ben Horowitz Says \$15 Million For Rap Genius Will Create ‚A Talmud For The Internet‘“, in: *FastCompany* (10. März 2012), <http://www.fastcompany.com/3001879/ben-horowitz-says-15-million-rap-genius-will-create-talmud-internet> (letzter Aufruf 24.7.2018).

¹⁴ So lautet das derzeitige Selbstverständnis auf *Genius*.

¹⁵ Andrea Lo (Anm. 10).

an. Medienökonomisch bedeutet dieser Vorgang, dass Rezipienten Teil der Kreativwirtschaft geworden sind. Für die Rap-Poeten verändert die Literarisierung ihrer performativen und oralen Kultur die bisherige Art zu texten. Ein zwar ob seiner Größe nicht zu überblickendes, aber elektronisch durchsuchbares Korpus ergänzt die ‚Quellen‘ des Rap und erhöht zugleich das intertextuelle Reflexionsniveau. Für Literaturhistoriker bedeutet die Verschriftung durch Amateure, d. h. Liebhaber bzw. Fans, das Aufkommen eines neuartigen literarischen Korpus: War Literatur zwar immer schon von ihrer Leserschaft her verstehbar, so waren doch mit ihrer Konstituierung und Herstellung vornehmlich Philologen und Verleger befasst. An *Genius* soll die neue kollektive Bildung von literaturgeschichtlich signifikanten Korpora vorgestellt und zugleich von kollektiven Praktiken der analogen Druckkultur abgegrenzt werden.

1 Für alle – von allen. Das populäre Lied-Korpus nach 1800 und nach 2000

Neue Gegenstände wissenschaftlicher Betrachtung stehen am Anfang der Disziplinengese. Objekte, für die kein Fach zuständig ist, werden zunächst als ein Liebhaberwissen kommuniziert, weshalb sich die initiale Motivation als eine erotische charakterisieren ließe. In dem Moment, in dem die Amateure auf Gleichgesinnte treffen, die in demselben Bereich Wissen begehren, entsteht ein sozialer und kommunikativer Raum als eine Wissensgemeinschaft. Die allmähliche Verfertigung der Disziplin beim Reden über ihre Gegenstände kann an der Germanistik beobachtet werden. Das Fach entstand in dem Moment, in dem aus Liebhabern mittelalterlicher Autoren und Werke zunächst Gelehrte und dann disziplinäre Fachleute wurden.¹⁶

Die Sammel- und Urteilsfreudigkeit philologischer Amateure, die gegenwärtig in den digitalen Räumen beobachtet werden kann, lässt sich mit der romantisch inspirierten durchaus vergleichen. Die Disziplinbildung gestaltet sich heute freilich anders, da das Spektrum der medialen Formen der Populärkultur weit mehr umfasst als populäre Lieder. Mit der Institutionalisierung der Medienkulturwissenschaft in Deutschland zeigt sich zudem, dass Philologie und Literaturwissenschaft nicht unbedingt zuständig sind. Die Vergleichbarkeit liegt also weniger im Prozess der Disziplinbildung und mehr in dem der Verschriftung. Nach 1800 und nach 2000 setzte jeweils ein internationaler Transkriptionsprozess¹⁷ ein, der bis dahin gesungene Lieder massiv in die schriftliche Kommunikation einspeiste.

¹⁶Vgl. Jürgen Fohrmann/Wilhelm Voßkamp (Hg.), *Von der gelehrten zur disziplinären Gemeinschaft*, Stuttgart 1987.

¹⁷Zu den deutschsprachigen Projekten zwischen 1770 und 1950 siehe Waltraud Linder-Beroud, *Von der Mündlichkeit zur Schriftlichkeit? Untersuchungen zur Interdependenz von Individualdichtung und Kollektivlied*, Frankfurt a.M. 1990.

Die Idee des Volksliedes oder der Sage ist die, dass ihre Überlieferung quasi von selbst stattfindet. Das Lied ist Volkslied, weil es vom Volk im Gesang tradiert wird. Es verbreitet sich in der Gemeinschaft idealerweise zwischen den Schichten und wird von Generation zu Generation fortgetragen. Philologen wie Johann Gottfried Herder oder Jacob Grimm waren als berufsmäßige Überlieferer fasziniert von der Vorstellung, dass die Überlieferung dem Anschein nach von selbst verläuft. Die Volksliedphilologen haben jedoch ihre eigene Rolle im Überlieferungsprozess kaum reflektiert. Ihre Verschriftung und Kommentierung hat zwangsläufig zur Literarisierung des Gegenstandes geführt und damit die kollektive nicht-schriftliche Überlieferung in Konkurrenz zu einer schriftlichen Überlieferung gestellt, die von Einzelnen oder kleinen Gruppen einer elitären Schriftgelehrsamkeit gesteuert wurde.

Der Unterschied zwischen beiden Epochen besteht u. a. darin, dass man als heutiger Beobachter populärer Tradierung die Verschriftung als konstitutiven Teil der Überlieferung anerkennen muss. Auch diesmal entsteht eine Konkurrenz zwischen nicht-schriftlicher und schriftlicher Tradierung des Liedes, doch erfolgt die Literarisierung nicht mehr durch wenige Schriftgelehrte. Die Lieder sind nicht mehr populär, weil sie von mündlichen Gemeinschaften tradiert würden, sondern in schriftbasierten Netzwerken, die von vielen Amateuren, Liebhabern bzw. Fans kontrolliert werden. Das Populäre hat den privilegierten Raum der Schriftgelehrten erfasst. Verbindendes Moment beider Epochen ist das Bewusstsein, der Kreis medialer Teilhabe würde sich entgrenzen.

Massenalphabetisierung, Industrialisierung des Buchdrucks, Leihbibliotheken und Periodika in Massenaufgaben hatten nach 1800 zu einer allmählichen Ausweitung des literarischen Mediums auf die gesamte Bevölkerung geführt. Unterhaltungsliteratur in auflagenstarken Periodika und die Entstehung der Populärkultur waren die Folge.¹⁸ Parallel dazu transformierte sich eine philologische Praxis, die zwar nicht auf die reale Gesamtleserschaft, aber doch auf ihr Idealbild abzielte. Herders globale Volkslieder¹⁹ waren der erste Stoß in diese Richtung; die Romantiker nach ihm beschränkten die populäre Trägerschaft stärker auf die Nation. Dadurch wurde die Erweiterung zugleich wieder räumlich begrenzt. Denn obgleich das Primärpublikum der Literatur vor 1800 ein kleines gelehrtes war, so war es doch universal

¹⁸Hügel lässt die Populärkultur mit der *Gartenlaube* 1853 beginnen. Ein wesentliches Kriterium der Populärkultur für Hügel ist Rezeptionsfreiheit: „Ohne Rezeptionsfreiheit, verstanden sowohl als Freiheit, das zu Rezipierende auszuwählen, als auch den Bedeutungs- und Anwendungsprozess mitzubestimmen – also ohne ein bestimmtes Maß an bürgerlichen Freiheiten –, gibt es keine Populäre Kultur.“ (Vgl. Hans-Otto Hügel, „Einführung“, in: Ders. (Hg.), *Handbuch populäre Kultur. Begriffe, Theorien und Diskussionen*, Stuttgart 2003, 1–22, hier: 6). Heute muss die Rezeptionsfreiheit um Urteilsfreiheit ergänzt werden.

¹⁹Dezidiert gegen die nationale Lesart von Herders *Volksliedern* vgl. Alexander Nebrig, „Die Welt als Lied. Der globale Anspruch von Herders Volksliedern“, in: Christian Moser/Linda Simonis (Hg.), *Figures des Globalen. Weltbezug und Welterzeugung in Literatur, Kunst und Medien*, Göttingen 2014, 315–325.

und kosmopolitisch orientiert.²⁰ Der gelehrte Anspruch, eine die Gelehrtenwelt transzendierende Leserschaft anzusprechen, konnte in einer Kultur attraktiv werden, die erkennt, dass zum einen nicht mehr ein kleiner Kreis von Gelehrten, sondern potenziell das ganze Volk Träger der Literatur ist, und dass zum anderen dieser Träger noch stumm und passiv ist. Der Philologe, der nun nicht mehr mit einer kleinen gelehrten Leserschaft rechnen durfte, sondern mit allen Lesern potenziell rechnen konnte, sprach im Namen derjenigen Gesamtgemeinschaft, die er selbst für mustergültig hielt. Die Phonozentristen des Sturm und Drang und der romantischen Philologie haben stark gemeinschaftlich gedacht und den Akzent auf das Populäre gelegt. Indem sie sich für dieses interessierten, projizierten sie den Wunsch nach Gemeinschaft, der aus ihrer Vereinzelung im kritischen Schriftraum entstanden war, ins Mündliche. Dabei konnten sie auf eine schichtenübergreifende Publikums-idee rechnen. Die Lied-Schrift war ein Kollektivmedium, das durch ein imaginäres Kollektiv gedeckt war.²¹ Wer Volkslieder bis ins 20. Jahrhundert sammelte und publizierte, tat dies im Interesse einer abstrakten Gemeinschaft (Region, Nation). Tatsächlich edierten einzelne Gelehrte die zahlreichen Zeugnisse, die meist sogar von denjenigen Personen, die die Volkslieder sangen, zugesandt worden waren. Dennoch hatte ihre Art der Partizipation eine andere Qualität als die gegenwärtige: Die analoge Partizipation war aufgrund der zentralen Stellung der Philologen auf dem Gebiet der Edition, Kritik und Hermeneutik und deren Nähe zum Verlagswesen asymmetrisch.

Nach 2000 dezentralisierte sich die Sammelpraxis. Die Passivität der Gesamtleserschaft ist längst nicht mehr der Normalzustand für Kreative. Das digitale Medium ist sowohl Rezeptions- als auch Gestaltungs- und Urteilsraum. Inhalte können zugleich gelesen, bereitgestellt und beurteilt werden. Es existiert jenseits der Forschungsinstitutionen eine philologische Kultur, die durch und durch populär ist, weil sie jedem die Möglichkeit der Partizipation bietet, vor allem aber, weil sie die Partizipation auf alle philologischen Praktiken ausgedehnt hat (Verschriftung, Edition, Kommentar, Kritik, Hermeneutik, Kanonisierung etc.). Zudem kann prinzipiell alles zum Gegenstand dieser Partizipation werden. Damit besteht die Populärkultur nicht mehr nur aus immateriellen Gütern, die konsumiert oder rezipiert werden, sondern zugleich auch aus einer populären Philologie, die den Güterverkehr durch Urteile begleitet.

²⁰ Für Escarpit verlief die soziale Ausweitung parallel zur Nationalisierung der Leserschaft: „Wie einst der Bürger des 14. und 15. Jahrhunderts das mönchische Lateinerei verhaftete Buch zur Sprache seines Alltags nötigte, so zwang nun im 19. Jahrhundert der neue Leser aus dem Volk sein nationales Idiom in das einst kosmopolitische Buch der Gebildeten“ (Robert Escarpit, *Die Revolution des Buches*, Gütersloh 1967, 20).

²¹ Zur Imagination des Sozialen bei Achim von Arnim vgl. Urs Büttner, *Poiesis des ‚Sozialen‘. Achim von Arnims frühe Poetik bis zur Heidelberger Romantik (1800–1808)*, Berlin 2015, 381–433.

Die Ausweitung des Populären vom Bereich der Rezeption auf den der Kritik unterscheidet die laienphilologischen *Scholars* der Gegenwart von den romantischen Gelehrten. Wenngleich sich die Romantiker für das Volk interessierten und es als Überlieferungsträger umwarben, waren sie doch selbst Teil der Gelehrtenrepublik. Daran änderte auch der Umstand nichts, dass viele von ihnen keine professionellen Wissenschaftler, sondern Amateure der deutschen Literatur waren, und dass bisweilen die philologische Praxis ‚wild‘ wirkte.²² Trotz allem kamen diese Liebhaber aus dem Innenraum der gelehrten Welt. Germanisten waren studierte Altphilologen, Juristen oder Theologen. Die Herstellung des Korpus war eine gelehrte Angelegenheit gewesen, auch wenn durch den Vorgang Volkslieder und Sagen dem Volk in schriftlicher Form dargebracht wurden.

Kollektivität wird von den neuen Amateuren des populären Liedes nicht mehr wie einst von Herder und den Romantikern für das Lied postuliert – ob nun in der oralen Entstehung oder durch mündliche Tradierung –, sondern die digitalen Datenbanknutzer verwirklichen sie bereits im Moment der Verschriftung. Obgleich die alten romantischen Sammler und Philologen gemeinschaftlich arbeiteten, stellten sie doch letztlich eine relativ kleine Gemeinschaft aus der literarischen, meist akademischen Elite dar. Sie waren aufgrund ihres hohen literaturgeschichtlichen und -kritischen Bildungsgrades keine typischen Träger der Populärkultur. Eher wirkten sie als ihre Agenten, wenn sie die Erzeugnisse der Populärkultur aufschrieben.

Die Schriftwerdung des Liedes ist folglich ein bereits geschichtlicher Vorgang, der einen ersten Höhepunkt in der Volksliedphilologie besaß, die – erinnert sei an die Volksliedstrophe und an Heinrich Heine – den Lyrikbegriff des 19. Jahrhunderts verändert hat. Es ist davon auszugehen, dass auch das neue Korpus seine Spuren in der Lyrik hinterlassen wird.²³ Auch wenn die meisten Lieder heute im Unterschied zu den alten Volksliedern autorisiert sind, ist das lyrikaffine Lied-Korpus der Gegenwart im Prozess seiner Literarisierung zu einem gemeinsamen Korpus geworden. Die Träger der populären Kultur prägen den Begriff des populären Liedes durch ihre kritische Verschriftung dergestalt um, dass er mit dem literaturwissenschaftlichen Gattungsbegriff der ‚Lyrik‘ kompatibel wird.

Nach dem typologischen Vergleich der beiden Lied-Korpora – Volkslied respektive populärkulturelles Lied – soll die jeweilige Form der Gemeinschaft rekonstruiert werden, die die Verschriftung verantwortet. Auch das alte literarisierte Volkslied war aus gemeinsamen Verschriftungsvorgängen entstanden. Aber zum einen bestehen qualitative und quantitative Unterschiede gegenüber dem

²²Vgl. Ulrich Wyss, *Die Wilde Philologie. Jacob Grimm und der Historismus*, München 1979, 282.

²³Studien zur Gegenwartslyrik, die das Rap-Korpus einbeziehen, werden künftig häufiger zu lesen sein, vgl. Maximilian Czollek, „Jubiljahre. Zur Aktualität theologischer Topoi in der deutschen Gegenwartslyrik“, in: Dörte Linke/Florian Priesemuth/Rosa Schinagl (Hg.), *Sprachen des Unsagbaren. Kulturelle Figurationen: Artefakte, Praktiken, Fiktionen*, Wiesbaden 2017, 243–266, hier: 247.

literarisierten Lied der Populärkultur, das aus der digitalen Medialität hervorgeht. Zum anderen zeigen Akteure der analogen Verschriftung um 1800 die Tendenz, sich als Agenten einer ausschließlich oralen Schöpfungs- und Überlieferungsgemeinschaft zu stilisieren.

2 Herders Erben im digitalen Raum

Die analogen Hersteller von Verschriftungen zu dem Zwecke, sie einem Publikum zu kommunizieren, hatten einen anderen Zugang zum populären Lied gefunden als die digitalen. Es sind nicht substanzielle Eigenschaften des jeweiligen Schriftmediums, die zu Unterschieden führen. Vielmehr verändert sich das Verhältnis von Mündlichkeit zur Schrift und dem Schreiben im Übergang von analoger zu digitaler Schriftlichkeit.

2.1 Die analoge Schriftgemeinde des Volksliedes

Der Schriftgelehrte Herder, der sich bereits im Alter von Mitte zwanzig mit einem Tintenfass verglichen hatte,²⁴ modellierte für seine 1778/79 erschienenen *Volkslieder* eine orale Schriftästhetik. Diese Schrift unternimmt alles, ihre Distanz zum Gesang abzubauen, unsichtbar zu werden und wie Gehörtes zu erscheinen. In Wirklichkeit ist die Distanz riesig, weil das eigentlich zeitlich erfahrbare Lied im uneigentlichen Medium des Schriftraums erfahren werden muss. Herder sagt zwar, dass das Lied „gehört werden“ müsse, „nicht gesehen“²⁵, täuscht aber darüber hinweg, dass seine Volkslieder für das hermeneutisch geschulte Auge von Lesern konzipiert sind: Herders Volkslieder sind eben keine Poesie für das Ohr.²⁶ Mündlichkeit zu fingieren, das zeigen auch die *Kinder- und Hausmärchen* der Brüder Grimm, wurde zu einer Hauptaufgabe einer Schrift im Dienste der mündlichen Volkskultur. Die um 1800 zu beobachtende Emphase auf die Stimme kann man somit auch kompensatorisch aus dem Fehlen phonographischer Techniken erklären.

Herder, der wie Goethe und die Generation des Sturm und Drang überhaupt ausgesprochen schriftfeindlich war, glaubte an das geschichtliche Primat des gesprochenen Wortes. Ursprünglich sei die orale Poesie gewesen, sekundär die schriftliche Literatur. Literatur verstand er als ein Phänomen gelehrter Bildung,

²⁴ Johann G. Herder, *Journal meiner Reise im Jahr 1769*, hg. von Katharina Mommsen, Stuttgart 1976, 9. Es handelt sich um den Beginn des postum erschienenen Reisejournals von 1769.

²⁵ Johann G. Herder, *Werke in zehn Bänden*, Bd. 3: *Volkslieder, Übertragungen, Dichtungen*, hg. von Ulrich Gaier, Frankfurt a.M. 1990, 247. Vgl. Nebrig (Anm. 19), 319–322.

²⁶ Vgl. Dieter Borchmeyer, „Poesie für das Ohr – Tönende versus gelesene Dichtung. Zur Geschichte eines Strukturproblems von Klopstock bis Wagner“, in: Walter Pape (Hg.), *Das „Wunderhorn“ und die Heidelberger Romantik. Mündlichkeit, Schriftlichkeit, Performanz*, Tübingen 2005, 207–223.

das seit der Renaissance die europäischen auf Mündlichkeit basierenden Poesiekulturen zerstört habe. Deshalb verglich er sich mit einem Tintenfass im polemischen Sinne Schillers, der das ganze Jahrhundert als tintenklecksendes Säkulum bezeichnete.²⁷

Herder bezog seine Motivation, Volkslieder zu sammeln und zu verschriftlichen, aus der Erfahrung übermäßiger Gelehrsamkeit. Die Schriftgelehrsamkeit erst bringt den Phonozentrismus hervor, und zwar bis zum performativen Widerspruch: Herders Quelle war nicht das gesungene Wort, sondern die Schrift: Abschrift, Umschrift von Liedern, weniger Niederschrift. Obgleich er also auf schriftlicher Vermittlungsebene ansetzte, betonte er den kollektiven Ursprung des Liedes, indem er vom Volkslied sprach. Anders als die gelehrte Kunstpoesie sollte mit wenigen Ausnahmen diese Form angeblich nicht mehr von einzelnen Dichtern geschaffen worden sein, sondern von Gemeinschaften.

Eng mit der Auffassung vom kollektiven Ursprung verbunden ist die Auffassung von der kollektiven Überlieferung. Genau genommen geht beides Hand in Hand. Wer versucht, sich einen kollektiven Ursprung vorzustellen, kann dies am besten dadurch, dass er sich die Mitglieder einer Gemeinschaft dabei vorstellt, wie sie das Lied oder die Sage von Generation zu Generation weitertragen. Die Tradition entsteht durch die Rezeptionsgemeinschaft bzw. die Gemeinschaft durch Rezeption gemeinsamer Lieder. Mit jeder Performanz wird das geistige Gut erweitert oder variiert, bis der Moment kommt, an dem es in die Schrift übergeht. Dass auch die Schrift Teil der Überlieferungskette mündlicher Poesie sein könnte, hat Herder ausgeblendet. Mit anderen Worten hat er die außerordentliche Rolle der Schrift für die Kanonisierung des Liedes verkannt. Für diese Verknennung ist auch der Märchensammler Jacob Grimm beispielhaft, der von einer Sage, die sich selbst ausspreche und verbreite,²⁸ fasziniert war. Wenn er denn Zeugen für seine Märchen anführen musste, dann erfolgte dies „pars pro toto“:²⁹ Eine Dorfbewohnerin repräsentiert die ganze Gemeinschaft.

Man kann den kollektiven Ursprung aber auch auf der Schriftebene ansetzen, und bereits um 1800 haben Liedsammler wie Friedrich David Gräter und Philologen wie Friedrich August Wolf daran erinnert:³⁰ Nicht das im Volk gesungene Wort, sondern seine Schriftwerdung ist der Ursprung und der Beginn der kollektiven Überlieferung. Mit der Verschriftung der Lieder sind diese ein neuer

²⁷ Friedrich Schiller, *Die Räuber. Ein Schauspiel*, Frankfurt a.M./Leipzig 1781, 17.

²⁸ Jacob Grimm, „Gedanken. Wie sich die Sagen zur Poesie und Geschichte verhalten“, in: *Zeitung für Einsiedler* (4./7. Juni 1808), 152–156, hier: 153.

²⁹ Heinz Rölleke, *Nebeninschriften. Brüder Grimm – Arnim und Brentano – Droste-Hülshoff. Literarhistorische Studien*, Bonn 1980, 1–15, hier: 2.

³⁰ Wolfs *Prolegomena ad Homerum* (1795) weist den Diaskeuasten eine zentrale Rolle bei der Strukturierung der homerischen Epen zu, vgl. Friedrich A. Wolf, *Prolegomena zu Homer*, übers. von Hermann Muchau, Leipzig 1908 (lat. 1795), 163 f. Zu Gräters Thematisierung der schriftlichen Überlieferung des Volksliedes in *Ueber die deutschen Volkslieder und ihre Musik* (1794) siehe Hermann Bausinger, „Gräters Beitrag zur Volksliedforschung“, in: *Jahrbuch Württembergisch Franken* 52 (1968), 73–94, hier: 88 f.

Gegenstand der Überlieferung geworden. Das homerische Epos, würde Wolf sagen, ist eine Erfindung von Philologen.

Die Gegenwart und sein eigenes Umfeld hätte Herder daran erinnern müssen, dass er in einem kleinen Kollektiv arbeitete: Freunde wie Goethe oder Familienangehörige wie seine Frau konnte er für das Volksliedprojekt gewinnen. Die postumen *Stimmen der Völker in Liedern* (1807) gingen nicht mehr auf das Konto Herders, womit der transgenerationale Charakter der das Volkslied herstellenden Gemeinschaft deutlich wird.³¹ In der Nachfolge Herders kann man sogar beobachten, dass nicht nur die Zahl der Liebhaber von Volksliedern, Sagen und Märchen zunahm, sondern dass diese immer häufiger zusammenarbeiteten. Sie kommunizierten im 19. Jahrhundert brieflich, organisierten sich in gelehrten Gesellschaften oder in philologischen Disziplinen.

Trotz dieser Hinweise auf Gemeinschaftsarbeit in der analogen Schriftkultur möchte ich den Unterschied zur gegenwärtigen kollektiven Arbeit auf keinen Fall nivellieren: Obzwar die Sammler des 19. Jahrhunderts zusammenarbeiteten, reden wir heute zu Recht von den Schöpfungen Einzelner: Herders Volkslieder, Brentanos und Arnims *Des Knaben Wunderhorn*, Grimms Märchen.³² Ein Grund dieser Individualphilologie des Populären ist in der Materialität der analogen Schriftlichkeit zu suchen, deren Verbreitung maßgeblich durch das Funktionieren der Beziehung des Aufschreibenden mit einem Verleger bestimmt war. Die Konzentration der Gemeinschaftsarbeit in einer als Herausgeber fungierenden Hauptperson ergab sich in der typographischen Epoche ganz von selbst.

2.2 Rahmenbedingungen digitaler Verschriftung

Mittlerweile gibt es phonographische Aufzeichnungsmedien, die die Schrift entlasten, Mündlichkeit zu simulieren, sobald das gesungene Wort Schrift wird. In der Geschichte der Lied-Transkription stellt das Auftreten der Schrift im digitalen Medienraum die wohl wichtigste Veränderung dar. Durch sie wurde die einst ubiquitäre Medialität der analogen Schrift zu einer Option degradiert, die zudem droht, antiquiert zu werden. Die analoge Hand- oder Typenschrift steht in Konkurrenz zu einer digitalen Schrift, die von Maschinen gelesen wird. Vier Konzepte – Transmedialität, Versatilität, Intertextualität und Interaktivität – halte ich für die Analyse digitaler Verschriftung für besonders kennzeichnend. Sie sind schon in der analogen Kultur des 20. Jahrhunderts theoretisiert worden und unmöglich nur auf das digitale Medium beschränkt. Jedoch eignen sie sich, die behauptete Neuartigkeit zu beschreiben. Transmedialität, Versatilität,

³¹Vgl. Kaspar Renner, „Des Vaters Wille“. Nachlassbewusstsein und Werkpolitik in der Familie Herder“, in: Kai Sina/Carlos Spoerhase (Hg.), *Nachlassbewusstsein. Literatur, Archiv, Philologie 1750–2000*, Göttingen 2017, 179–216.

³²Die Rechtstheorie, namentlich die von Vesque von Püttlingen (*Das musicalische Autorrecht* [1864]), hat bereits im 19. Jahrhundert den Sammler mit Autorrechten versehen, vgl. Monika Dommann, *Autoren und Apparate. Die Geschichte des Copyrights im Medienwandel*, Frankfurt a.M. 2014, 36–41.

Intertextualität und Interaktivität lassen sich im Sinne von Jay D. Bolter als Qualitäten der digitalen Schrift verstehen, die durch die digitale Technik weder erfunden worden seien noch die Kultur der Gegenwart determinierten. Sie gehörten aber zu jenen Weisen des Schreibens, die die Computertechnologie einfacher gemacht habe als andere.³³

Das stärkere Hervortreten jener vier Qualitäten in digitaler Schrift ist Folge des dynamischen Verhältnisses der drei Ebenen, auf welchen digitale Schrift zugleich realisiert werden muss: der *binären Ebene*, der *Auszeichnungsebene* und der *Oberflächenebene*. Die *binäre Ebene* erinnert daran, dass sich Schrift im Digitalen prinzipiell nicht von anderen Medien unterscheidet. Sobald eine Information lesbar für die Maschine werden soll, muss sie in eine Abfolge von Nullen und Einsen überführt werden. Lesbarkeit meint hier den Rhythmus von ‚An‘ und ‚Aus‘. Ob ein Bild, ein Schriftbild, ein Ton, ein Schriftton oder sonst etwas digitalisiert vorliegt, ist nicht mehr unterscheidbar auf dieser Ebene. Das bedeutet, dass Schrift hier nicht mehr existiert oder immer schon einen transmedialen Charakter besitzt. Die *Auszeichnungsebene* ist in der analogen Druckkultur zwar vorhanden. Wie die Schriftgestalt auszusehen hat: kursiv, recte, gesperrt, im Blocksatz usw., ist vorher festzulegen. Jedoch ist die Auszeichnung fix bzw. kann nur verändert werden um den Preis eines Neudrucks. Auszeichnungspraktiken sind kostspielig, einmalig und unwiderruflich. Volksliededitionen beispielsweise müssen vorher auszeichnungstechnisch genau konzipiert werden. Der Text ist identisch mit seiner Materialität, von ihr nicht abziehbar und druckhistorisch in Auflagen pluralisiert. Die digital durch eine *Markup Language* ausgezeichnete Schrift dagegen ist dynamisch, veränderbar, immateriell, auszeichnenbar für verschiedene Endgeräte und Leserwünsche.³⁴ Die Kopplung an die Auszeichnungsebene hat wiederum Folgen für die *Oberflächenebene*. Ein mit sich identisches Schriftbild gibt es nicht bzw. lässt es sich wie in der PDF nur vortäuschen. Transmedialität, Versatilität, Intertextualität und Interaktivität können auch als Phänomene schriftlicher Performanz verstanden werden.³⁵ Die Betreiber der Datenbank zur Verschriftung und Kommentierung von Liedern *Genius* haben sich dafür entschieden, den Prozess der Verschriftung durch diese vierdimensionale Performanz zu gestalten. Diese Offenheit beispielsweise unterscheidet sie von einer Plattform wie *songtexte.com*, die die Texte nur lesbar macht.

³³Jay D. Bolter, „Das Internet in der Geschichte der Technologien des Schreibens“, in: Sandro Zanetti (Hg.), *Schreiben als Kulturtechnik. Grundlagentexte*, Frankfurt a.M. 2015, 318–337, hier: 320. Vgl. zur digitalen Schrift auch ders., „Digitale Schrift“, in: Gernot Grube/Werner Kogge/Sybille Krämer (Hg.), *Schrift. Kulturtechnik zwischen Auge, Hand und Maschine*, München 2005, 453–467.

³⁴Axel Kuhn/Svenja Hagenhoff, „Digitale Lesemedien“, in: Ursula Rautenberg/Ute Schneider (Hg.), *Lesen. Ein interdisziplinäres Handbuch*, Berlin 2015, 361–380, hier: 365–371.

³⁵Vgl. bereits Uwe Wirth, „Performative Rahmung, parergonale Indexikalität. Verknüpfendes Schreiben zwischen Herausgeberschaft und Hypertextualität“, in: Ders. (Hg.), *Performanz. Von der Sprachphilosophie zu den Kulturwissenschaften*, Frankfurt a.M. 2002, 403–433.

- a) Der *transmediale* Charakter führt dazu, dass der Schriftraum potenziell ein audiovisueller Raum ist. Anders als in den *Volksliedern* Herders, der nicht einmal Noten abdruckte, ist den neuen literarisierten Liedern sogar eine mündliche Form der Performanz beigegeben. Der Raum des Kommentars arbeitet nicht nur mit Schrift, sondern auch mit der Evidenz der Bilder, Videos und Töne. Das Kommentieren von Texten war in den analogen Handschriften- und Druckepochen eine Praxis, die sich auf die Schrift beschränkte und kostenbedingt sehr selten auf Bildmedien und Notenzeichen zurückgriff. Nunmehr ist die Schrift nicht mehr das Hauptmedium der Hermeneutik. Die Verschriftung erfolgt in einem semi-oralen Medium. Epistemische Gewinne werden dort verbucht, wo das begleitende Tondokument etwa aufschlussreich für die Lautstruktur ist.³⁶
- b) Der *versatile* Charakter der digitalen Schrift zeigt sich darin, dass es weder ein festes Schriftbild noch ein stabiles paratextuelles Seitenlayout gibt. Herders *Volkslieder* beispielsweise erschienen in der Fraktur, ebenso lange Zeit auch Grimms Märchen. *Genius* verwendet eine serifenlose Schrift, aber diese Wahl ist immer nur vorläufig bis zur nächsten Reform des Seitenlayouts. Die Dynamik der Schrift wird auch über ihre Paratexte deutlich, die als Werbung und Seitengestaltung permanent differieren. Je nach Nutzerinteresse, kommerzieller Natur des Anbieters und Endgerät sehen das, was wir lesen, und das, worin es eingebettet ist, anders aus.³⁷ Der bewegte Text, der für die Dekonstruktion noch ein Phantasma war, ist nunmehr wirklich in Bewegung geraten.³⁸ Der dynamische Text der Poststrukturalisten war jedoch vor dem Hintergrund eines stabilen Textmediums mit einem festen Satzspiegel und unverrückbarer Typographie

³⁶ Einige Lieder wie *Rap God* von Eminem enthalten im Kommentarteil eine Analyse der lautlichen Wiederholungsstruktur (Reime, Assonanzen). Dabei wird der Text parallel zur Musik nach und nach sichtbar gemacht. Wiederholungen werden farblich markiert. Vgl. <https://genius.com/Eminem-rap-god-lyrics> (letzter Aufruf 24.7.2018).

³⁷ Zur Paratextualität im Netz siehe Annika Rockenberger, „Paratext‘ und Neue Medien. Probleme und Perspektiven eines Begriffstransfers“, in: *Philologie im Netz* 76 (2016), 20–60.

³⁸ Bolter 2015 (Anm. 33), 323, meint, dass für das Phänomen des digitalen Schreibens in Ermangelung eines Besseren die „Rhetorik des Poststrukturalismus“ verwendet werde. Dem Zusammenhang zwischen Literaturtheorie und den technischen Vernetzungsmöglichkeiten des digitalen Mediums widmet sich Landow 2006 (George P. Landow, *Hypertext 3.0. Critical Theory and New Media in a Global Era*, Baltimore 2006 [1992]).

Als Beispiel dafür, dass die poststrukturalistische Theorie die Text- und Schriftpraxis der Gegenwart antizipiert hat, lese man das Referat über den poststrukturalistischen Textbegriff von Eagleton (Terry Eagleton, *Einführung in die Literaturtheorie*, Stuttgart 2012 [zuerst dt. 1988], hier: 103 u. 106).

Die Faszination poststrukturalistischer Denker für bewegte Texte und ihre Polemik gegen feste Werke können als theoretische Reflexe auf den Zusammenschluss von Elektronikkonzernen und literarischen Medienunternehmen begriffen werden. So kaufte die *Radio Corporation of America (RCA)* 1965 den Verlag *Random House* auf. Polemisch, wenngleich erhellend, beschreibt die Vorgänge Schiffrin (André Schiffrin, *Verlage ohne Verleger. Über die Zukunft der Bücher*, übers. von Gerd Burger, Berlin 2000, 42–44).

imaginiert worden. Denn obwohl die Überlieferung in verschiedenen Editionen, Übersetzungen, Buchgestaltungen pluralisiert vorlag, las man im analogen Zeitalter gewöhnlich eine einzelne Ausgabe, außer man verglich gerade kritisch zwei Ausgaben miteinander. Die Identität des Textes wurde im analogen und gedruckten Buch garantiert, und konkurrierende Ausgaben irritierten nicht wirklich das Lektüerverhalten. Bis zum Ende des 20. Jahrhunderts waren Lese-medien „durch ihre Materialität und statische Eigenschaften geprägt“.³⁹ Im Digitalmedium jedoch ist „die Einheit von Zeichencodierung und materiellem Objekt aufgelöst“.⁴⁰ Es gibt folglich keine vollendeten Texte mehr: „Das Werk stellt kein geschlossenes Objekt dar, sondern es erscheint vielmehr als eine prozessierende Relation zwischen Code und Darstellung.“⁴¹ Statt von einer festen Form der Textoberfläche ist von einer *versatilen* Form als Normalfall auszugehen, veränderbar auf Auszeichnungsebene, abgestimmt auf Endgeräte und Nutzervorlieben.

- c) Der *intertextuelle* Charakter digitaler Schrift resultiert aus der *vertikalen Vernetzung* mit anderen Schriftformen und Medien außerhalb des primären Schrift-raumes. Gemeint ist das, was Kristeva als ‚Intertextualität‘ konzipiert hatte: der Bezug des Wortes im Text zu anderen Texten und Wörtern entweder der Gegenwart oder der Vergangenheit.⁴² Jedes Element einer Schrift verweist auf ein anderes. Der Verweisungszusammenhang ist unbegrenzt: nicht nur schrift-intern, sondern auch intermedial. Es obliegt dem Leser, den Kontext als den mitzulesenden Text zu definieren. Textausgaben des Druckzeitalters haben in Kommentaren versucht, die Prä- und Intertexte einzubeziehen. Sie konnten den Außenbezug des Wortes oder des ganzen Textes jedoch nicht vollständig abbilden, weil die Ausgabe irgendwann fixiert und abgeschlossen werden musste. Die Vernetzung der digitalen Schrift ist prinzipieller. Wissen über Textbezüge kann in größeren Mengen akkumuliert und schneller sichtbar gemacht werden. Im Medium der digitalen Schrift lässt sich die Geschlossenheit des Werks schwerer behaupten und das Korpus nicht mehr begrenzen. Auch intermediale Bezüge können problemlos sichtbar gemacht werden, was beispielsweise eine Sammlung von Filmziten in Liedern belegt.⁴³ Der kommunikative Charakter der oralen Poesie wird ebenfalls evident. Wie in der Lyrik die Dichter im Gespräch mit anderen Dichtern stehen, so kommunizieren auch im Rap die Vortragskünstler

³⁹ Kuhn/Hagenhoff (Anm. 34), 362.

⁴⁰ Ebd.

⁴¹ Jörgen Schäfer, „Netzliteratur“, in: Natalie Binzcek/Till Dembeck/Jörgen Schäfer (Hg.), *Medien der Literatur. Ein Handbuch*, Berlin 2013, 481–501, hier: 485.

⁴² „Le statut du mot se définit [...] *verticalement*: le mot dans le texte est orienté vers le corpus littéraire antérieur ou synchronique.“ Vgl. Julia Kristeva, „Bakhtine, le mot, le dialogue et le roman“, in: *Critique. Revue générale des publications françaises et étrangères* 23 (1967), 438–465, hier: 440.

⁴³ <https://genius.com/Genius-deutschland-samples-und-zitate-aus-film-und-fernsehen-annotatedb> (letzter Aufruf 24.7.2018).

über ihre Lieder intensiv miteinander. Eine belanglos anmutende Zeile wie „Stress ohne Grund“⁴⁴ wird im digitalen Editionsraum als ein epistemisches Element erkennbar, das den kommunikativen Zusammenhang belegt. Um die Dimension der produktiven Rezeption einzufangen, sei auf die Möglichkeit verwiesen, Parodien, Kontrafakturen, Übersetzungen und andere Formen der Nachahmung in die Darstellung einzubeziehen: Siebzig Cover-Versionen konnten bislang allein zu *Blowin' in the Wind*⁴⁵ zusammengetragen werden.

- d) Der *interaktive* Charakter schließlich resultiert aus der *horizontalen Vernetzung* des Wortes zum Publikum im Sinne Kristevas.⁴⁶ Der horizontale Bezug der digitalen Schrift als der Bezug zum Rezipienten ist nicht mehr der analoge einer homogenen Leserschaft. Im Digitalen wird die homogene Gruppe in ihre einzelnen Leser pluralisiert. Digitale Schrift ist bei *Genius* kollektiv. Vor allem aber ist der Schriftempfänger potenzieller Schriftsender. Alles kann von jedem kommentiert werden. Das kollaborative und interaktive Potenzial der Digital-schrift wird von der Wissensdatenbank genutzt – mit Auswirkungen auf die Wissensproduktion. Im Unterschied zur analogen Tradition des Volksliedes ist nunmehr Verschriftung ein explizites Gemeinschaftsprojekt. Denn Überlieferung wird nicht mehr nur im Raum mündlicher Performanz von einer populären Rezeptionsgemeinschaft bewerkstelligt, sondern eben diese ‚elektronische Gemeinschaft‘⁴⁷ oder wenigstens ein Teil von ihr setzt im Raum der Schrift die Überlieferung fort. War populäre Überlieferung des Liedes in der analogen Epoche ein Phänomen der Mündlichkeit und ihrer Performanz, so ist in den letzten Jahren das Lied auch in der Schrift ‚volkläufig‘⁴⁸ (John Meier) geworden.

3 Kollektive Überlieferung auf *Genius*

Transkription und Kommentierung auf *Genius* lassen sich an sehr heterogenen textuellen Formen und Gattungen beobachten. Prinzipiell kann jeder Text eingestellt werden, sodass oft bereits textbasierte Inhalte hineinkopiert wurden, die nicht zum transkribierten Korpus gehören. Zu Liedern und sangbaren literarischen Texten gehören auch historische Gedichte, Volkslieder oder Songs, wobei sich die

⁴⁴Ali Bumaye, „Sex ohne Grund“ (Kommentar zur Zeile „Stress ohne Grund“: <https://genius.com/9135027> [letzter Aufruf 24.7.2018]).

⁴⁵<https://genius.com/Bob-dylan-blowin-in-the-wind-lyrics> (letzter Aufruf 24.7.2018).

⁴⁶„Le statut du mot se définit alors a) *horizontalement*: le mot dans le texte appartient à la fois au sujet de l'écriture et au destinataire“ (Kristeva [Anm. 42], 440).

⁴⁷Zu diesem Konzept siehe Bolter 2015 (Anm. 33), 331–334.

⁴⁸John Meier, „Einführung“, in: Ders. (Hg.), *Das deutsche Volkslied*, Bd. 1., Leipzig 1935, 7–34, hier: 7.

Aufmerksamkeit derzeit hauptsächlich auf die angelsächsische Literatur oder aber auf die Weltliteratur in englischer Übersetzung richtet.⁴⁹

Nicht jedoch die literarische Tradition, sondern die Lieder der jüngeren Populärkultur, insbesondere des Rap, erregen auf der Plattform das hauptsächliche Interesse. *Genius* ist primär kein wissenschaftliches Beobachtungsmedium. Das Netzwerk dient ebenso der Selbstverständigung von Sängern und Hörschaft. Ein Lied wie *Rap God* von Eminem zeigt das quantitative Potenzial: 14 *Scholars* sind an der Niederschrift beteiligt gewesen und 1.379 (1.800 [6/2022]) weitere haben es kommentiert, darunter der Künstler selbst. Insgesamt wurden 118 Kommentare durch Moderatoren akzeptiert (*Accepted Annotations*) und drei vom Künstler geprüft (*Verified Annotations*). Über 12,5 Mio. Mal (8/2018; 17,7 Mio [6/2022]) wurde der Text aufgerufen.⁵⁰ Anschlusskommunikation durch die Schrift wird ermöglicht und Wissen über verschiedene populärkulturelle Formen verzeichnet, wobei die *Wikipedia* als Wissensquelle sowie als anschließendes Kanonisierungsmedium von zentraler Bedeutung ist.

Zu den populären deutschen Liedern gehört *Chabos wissen wer der Babo ist* (2012) von *Haftbefehl*,⁵¹ das in professionellen bzw. redaktionellen Medien wie Tageszeitungen und Fernsehsendungen größere Bekanntheit erreichte.⁵² Diese bestätigen auch die Parodien.⁵³ Genauer betrachtet werden soll das laienphilologische Verhalten der Netzwerknutzer jedoch nicht an diesem Lied, sondern an dem parallel erschienenen Remix, der insgesamt zwölf Künstler vereint und deutlich umfangreicher ist. Die sprachliche Komplexität ergibt sich aus der Mehrsprachigkeit des Liedes. Die kritische Überlieferung im engeren Sinn fand sowohl auf YouTube durch den Künstler selbst⁵⁴ als auch auf *Genius* statt. Sechs

⁴⁹ Shakespeares *Sonett 18* beispielsweise wurde von 45 *Scholars* kommentiert und 97.800 Mal aufgerufen. *Auld Lang Syne* nach dem Text von Robert Burns haben zehn *Scholars* kommentiert; verzeichnet sind knapp 13.000 Aufrufe (Stand 7/2018). Bob Dylans *Blowin' in the wind* hat 48 *Scholars* zur Kommentierung veranlasst und erhielt 224.800 Aufrufe. Dabei kamen innerhalb eines Jahres (von Juli 2017 bis Juli 2018) sechs Kommentatoren und 8.700 Leser des Sonetts hinzu. Bob Dylans Lied wurde deutlich stärker rezipiert: Zehn weitere Kommentatoren und 130.000 neue Aufrufe kamen im selben Zeitraum hinzu. Bei den drei Beispielen handelt es sich um kanonische Texte, für die auch *Wikipedia*-Einträge in verschiedenen Sprachen vorliegen. Vgl. <https://genius.com/William-shakespeare-sonnet-18-shall-i-compare-thee-to-a-summers-day-annotated>, <https://genius.com/Robert-burns-auld-lang-syne-annotated>, <https://genius.com/Bob-dylan-blowin-in-the-wind-lyrics> (letzter Aufruf der Links 24.7.2018).

⁵⁰ <https://genius.com/Eminem-rap-god-lyrics> (letzter Aufruf 24.7.2018).

⁵¹ Bürgerlich Aykut Anhan, der das Lied in Zusammenarbeit mit Farid Bang (Farid El Abdellaoui) performiert hat.

⁵² Daniel Haas, „Chabos wissen, wer der Babo ist. Gaumenfreuden aus dem Getto. Der Rapper Haftbefehl macht Metasprache aus prekärem Slang“, in: *FAZ.NET* (31. Januar 2013), <http://www.faz.net/aktuell/feuilleton/pop/rapper-haftbefehl-chabos-wissen-wer-der-babo-ist-12046385.html> (letzter Aufruf 24.7.2018).

⁵³ https://www.youtube.com/watch?v=_bL2iKbaavU (letzter Aufruf 24.7.2018).

⁵⁴ Interview auf *Clixroom* am 17.1.2013: https://www.youtube.com/watch?v=h9Q_gE2MhKk (letzter Aufruf 24.7.2018).

Scholars haben es ediert, und bis Juli 2018 kommentierten es 103 weitere. Diese kommentierte Edition wurde bis Juli 2017 von 545.000 Menschen gelesen.⁵⁵ Bis Juli 2018 wurden es 567.600 (618.600 [6/2022]). Bevor anhand dieses Beispiels die laienphilologische Praxis auf *Genius* analysiert wird, seien kurz die Nutzungsbedingungen und Nutzerrollen vorgestellt.

3.1 Bedingungen der Partizipation auf *Genius*

Neben dem einfachen Beiträger (*Contributor*) gibt es den Editor (*Editor*), den Vermittler (*Mediator*) und den Moderator (*Moderator*). Eine vierte Gruppe bilden geprüfte Künstler (*Verified Artists*), die mit erheblich mehr Grundautorität den diskursiven Raum betreten.

Um als Nicht-Künstler eine der drei Rollen zu erhalten, muss man sich zuvor durch ein Verhalten qualifiziert haben, das der Wissensproduktion zugutekommt. Für jeden Kommentar (*Annotation*)⁵⁶ erhält man fünf sogenannte ‚IQ-Punkte‘, die symbolische Währung der Plattform. Überschreitet man die 300, wird man als Editor, Mediator oder Moderator zugelassen, kann sich also entscheiden, ob man vorzugsweise transkribiert und editiert, lieber die Kommunikation innerhalb der Plattform fördert oder aber strukturelle Verantwortung übernimmt. Der Gründer der deutschen Abteilung von *Genius*, Tobias Wilinski, ist folglich Moderator. Insgesamt gibt es vier Moderatoren (6/2022: 9), aber nur einen Mediator (6/2022: 2) für die deutsche Kommunikation innerhalb der Liste, deren Akteure teilweise inaktiv sind. Die meisten Beiträger werden Editoren, die für die Qualitätskontrolle sorgen, wobei zahlreiche unbearbeitete (*unreviewed*) Beiträge auf Kontrolle warten.

Wird der Zeilenkommentar eines Beiträgers durch den Editor zugelassen, erhält man weitere zehn Punkte; eine Ablehnung führt zum Verlust von sieben Punkten. Gute Kommentare im Sinne von *Genius* erläutern Referenzen zu Songs und Filmen, definieren Slang-Ausdrücke, übersetzen Fremdwörter, kontextualisieren Zeilen durch Interviews, verweisen auf geschichtliche und gegenwärtige Ereignisse, argumentieren biographisch („Verbindungen zum echten Leben des Künstlers“⁵⁷), beziehen – und das unterscheidet die digitale Annotation von der traditionellen Praxis – Bilder und GIFs ein. Ein guter Kommentar könne auch Witze „mit wertvollen Informationen“⁵⁸ enthalten.

Die Annotationsethik in zehn ‚Geboten‘ legt Wert darauf, Wiederholungen zu vermeiden, den Text im Kommentar nur zu paraphrasieren (1). Weiter wird auf eine einfache Sprache geachtet – „Schreib, als würdest du mit einem Freund

⁵⁵ <https://genius.com/Haftbefehl-chabos-wissen-wer-der-babo-ist-remix-lyrics> (letzter Aufruf 24.7.2018).

⁵⁶ Die deutsche Übersetzung ‚Annotation‘ wird auf *Genius* gleichbedeutend mit ‚Kommentar‘ benutzt.

⁵⁷ <https://genius.com/9074554> (letzter Aufruf 24.7.2018).

⁵⁸ Ebd.

über den Song reden“⁵⁹ und nicht wie ein „Roboter“ (2) –, auf Orthographie und Grammatik (3), auf Recherche, Belege und das Vermeiden von Spekulationen, kurz auf Redlichkeit (4), auf Sachlichkeit (5), auf eine gewisse Ökonomie der Erklärung (6), auf zeitliche und sachliche Genauigkeit (7, 8), auf die richtige Formatierung (9). Schließlich wird betont, syntagmatische Zusammenhänge statt einzelner Wörter zu kommentieren (10).

Richtungsweisend für die Transkription ist das Ethos der selbstständigen Transkription: „Geh nicht davon aus, dass die Lyrics auf anderen Lyricseiten richtig sind. Hör dir das Lied immer selbst an.“⁶⁰ Das Primat des Gehörten soll davon abhalten, die Texte aus dem Booklet abzuschreiben.⁶¹ Unverständliche Stellen werden durch ein Fragezeichen in eckigen Klammern ‚[?]‘ sichtbar gemacht. Alles soll notiert werden, also auch wiederkehrende Elemente wie der Refrain. Der Grundsatz, eine eigenständige Transkription zu erstellen, hat keine Entsprechung im Bereich der Literatur: Autopsie ist nicht vorgesehen, Textkritik von marginaler Bedeutung. Welcher Ausgabe beispielsweise der Moderator Tobias Wilinski (238.152 IQ) den *Faust*-Text entnimmt, wird nicht erwähnt.⁶²

Obleich jede Transkription auf dem persönlichen Gehör basiert, ist das Erscheinungsbild der Texte relativ homogen. Dafür sorgen Formatierungsregeln, die die Transkription organisieren.⁶³ Orthographie und grammatische Richtigkeit führen zur Normierung der sprachlich oftmals eigenwilligen Lieder; deutlich wird zudem, dass die Verschriftung keine Transliteration ist, was dann wiederum dazu führt, im Kommentar besondere Aussprachephänomene erläutern zu müssen.⁶⁴ Wie problematisch eine solche Standardisierung der Oralität ist, wird an Liedern deutlich, die ganz offensichtlich dialektal sind. Zusätzlich hat *Genius* deshalb einen Leitfaden für die Verschriftung von österreichischen Liedern, aber auch für den Sprachgebrauch im Kommentarbereich eingeführt.⁶⁵ In der Zeichensetzung werden Kommata regulär verwendet, auf Punkte soll weitgehend verzichtet werden. Die korrekte Apostrophierung wird ebenfalls erläutert. Die Zeilenformatierung ist zentral für das schriftliche Erscheinungsbild des Liedes. Zeilenumbrüche finden

⁵⁹ Ebd.

⁶⁰ <https://genius.com/9077261> (letzter Aufruf 24.7.2018).

⁶¹ „Schreibe genau das auf, was du auch hörst! Manchmal kann man Songtexte im Booklet eines Albums nachlesen oder bekommt die Texte aus anderen Quellen. Es kann aber sein, dass der Künstler noch kleine Änderungen vorgenommen hat und das Endergebnis leicht vom Booklet abweicht. In so einem Fall solltest du immer das transkribieren, was du im Song hörst, auch wenn im Booklet etwas anderes steht.“ (<https://genius.com/10917698> [letzter Aufruf 24.7.2018]).

⁶² <https://genius.com/Johann-wolfgang-von-goethe-faust-zueignung-kapitel-1-annotated> (letzter Aufruf 24.7.2018).

⁶³ <https://genius.com/10913272> (letzter Aufruf 24.7.2018).

⁶⁴ „Attention ist französisch und englisch für Achtung! (Er betont es allerdings ziemlich französisch.)“ (<https://genius.com/1311150> [letzter Aufruf 24.7.2018]).

⁶⁵ <https://genius.com/Genius-osterreich-osterreichischer-transkriptionsguide-annotated>, <https://genius.com/11.020.775>, <https://genius.com/11020840> (letzter Aufruf der Links 24.7.2018).

taktbasiert statt, aber auch Pausen sollen beachtet werden. Für die Gliederung werden Strophen oder Parts eingeführt, der Refrain soll als ‚Hook‘ bezeichnet werden. Die Markierung garantiert, den Text auch optisch als Song und nicht als Gedicht zu verstehen. Die Einbeziehung sogenannter *Ad-libs* (Sprache im Hintergrund) führt im zweidimensionalen Schriftbild zu einer gleichgeordneten Sprachebene, die die literarische Hermeneutik tangiert.

Eine weitere Möglichkeit, die epistemische Währung ‚IQ‘ zu akkumulieren, besteht im Schreiben von Biogrammen zu Künstlern sowie in der Zusammenfassung von Songs oder Alben. Durch das lobende Urteil einer Anmerkung kann man ebenfalls Punkte sammeln, wobei je nach Autorität zwei oder vier Punkte (ab 1.000 IQ) vergeben werden. Da alle Nutzer Teil eines Netzwerkes sind, kann man ihnen auch folgen. Das Teilen einer Anmerkung wird ebenfalls honoriert (*Pyongen*) sowie das Hinzufügen von Verbesserungsvorschlägen.

3.2 Populäre Philologie am Beispiel von *Chabos wissen wer der Babo ist*

Der spielerische Charakter⁶⁶ und die Möglichkeit, symbolisches Kapital zu erwerben, sind Strukturmaßnahmen, um das intrinsische Interesse am Gegenstand längerfristig aufrechtzuerhalten, die Qualität zu verbessern und die Aktivität der Plattform zu sichern. Transkription und Kommentar seien am Beispiel von *Chabos wissen wer der Babo ist* (Remix) nun genauer erläutert.

Die Transkription dieses Liedes erfolgte zu 78 % durch Hylia (Editor mit 55.020 IQ), zehn Prozent steuerte Bemoty (Editor mit 115.469 IQ) bei. Der Rest von zwölf Prozent verteilt sich auf zwei Beiträger (*Contributor*), einen Editor und den Moderator Tobias Wilinski. Angezeigt wird nur der aktuelle epistemische Wert, nicht der Wert zum Zeitpunkt der Transkription.

Unter den 103 Kommentatoren nimmt Hylia, hauptverantwortlich für die Transkription, die fünfte Stelle ein (213 IQ im Rahmen der Annotation), Tobias Wilinski Platz zwei (435). Die meisten Annotationen stammen von rayo840 (Beiträger mit 12.830 IQ, davon 909 IQ-Punkte für die Lied-Annotation zum Remix *Chabos wissen wer der Babos ist*). Das Rap-Duo Celo & Abdi ist an vierter Stelle, das sich zweimal mit einer Tonaufnahme in den Kommentar einschaltet.⁶⁷ Im Remix bestreiten die ‚geprüften Künstler‘ die Parts fünf und sechs des Liedes.

Im ‚Hook‘ oder Refrain des Remix werden fünf Stellen als kommentierungsbedürftig ausgewiesen: 1. „Chabos wissen, wer der Babo ist“, 2. „Hafti Abi ist der“, 3. „der im Lambo und Ferrari sitzt/Saudi Arabi Money Rich“, 4. „Attention,

⁶⁶Maina und Suleman diskutieren *Gamification* als Methode zur Erschließung des digitalen Erbes auch mit Bezug auf *Genius*, vgl. Job King'ori Maina/Hussein Suleman, „Enhancing Digital Heritage Archives Using Gamified Annotations“, in: Robert B. Allen/Jane Hunter (Hg.), *Digital Libraries: Providing Quality Information*, Cham 2015, 169–179, hier: 169 f.

⁶⁷<https://genius.com/1311251>, <https://genius.com/1344729> (letzter Aufruf der Links 24.7.2018).

mach bloß keine Harekets/Bevor ich komm' und dir deine Nase brech'", 5. „Wie er grade Nasen snifft“. Die Praxis wird kurz an der ersten Zeile erläutert, wobei hier auf den Kommentar zum Remix und zum Original zugleich zu sehen ist:

Chabos wissen, wer der Babo ist
 Hafti Abi ist der, der im Lambo und Ferrari sitzt
 Saudi Arabi Money Rich
 Wissen, wer der Babo ist
 Attention, mach bloß keine Harekets
 Bevor ich komm' und dir deine Nase brech'
 Wissen, wer der Babo ist
 Immer noch der selbe Chabo, Bitch
 Den du am Bahnhof triffst, wie er grade Nasen snifft
 Wissen, wer der Babo ist
 W-W-Wissen, wer der Babo ist⁶⁸

Als Erstkommentar zur ersten Zeile des Originallieds stand: „Babo ist eine andere Form von ‚Baba‘ und bedeutet Vater. In diesem Fall ist es eine Machtdemonstration.“⁶⁹ Der Satz datiert aus dem Jahr 2012 (Beiträger: agitdersim, 724 IQ). Ein Jahr später, 2013, ergänzt der Editor 13XxM1CHA3lxX37 (121.198 IQ) einen Verweis auf die Remix-Strophe, in der es heißt: „wir Berliner, sagen zu Chabos Bruder, zu Babo Boss“ sowie auf einen stützenden *Welt*-Artikel gleichsam als Autoritätsargument: „Diese Definition unterstützt auch ein Artikel auf welt.de, in dem es heißt: ‚Chabo‘ ist ein Wort aus der mittelalterlichen Gaunersprache Rotwelsch und bedeutet ‚Junge‘. ‚Babo‘ kommt aus dem Türkischen und heißt ‚Chef‘ oder ‚Boss‘.“⁷⁰ Der erwähnte Artikel datiert vom 2.5.2013⁷¹ und unterstreicht nochmals, dass andere Schriftmedien frühzeitig am hermeneutischen Prozess und an der kritischen Überlieferung beteiligt gewesen sind. Es folgen noch vier Änderungen im Kommentar, die aber nur die Orthographie betreffen (‚ehr‘ zu ‚eher‘). Argumentiert wird für ‚Boss‘ statt ‚Vater‘, d. h. mit Verweis auf eine Parallelstelle im Remix sowie mit einem Beleg aus einem anerkannten Zeitungsmedium, dem Feuilleton der *Welt*. Im Remix hatte die Erstkommentierung ebenfalls schon 2012 eingesetzt, aber bis 2014 konnte man nur den Satz lesen: „Chabo bedeutet auf Romanes, also auf Zigeunerisch, Junge, und wird oft in Frankfurt verwendet“ (Beiträger: rayo840, 12.830 IQ).⁷² Eine renommierte Editorin der Plattform, SinaTheQueen (Editor, 232.442 IQ), begann 2014, den Kommentar zu

⁶⁸ <https://genius.com/Haftbefehl-chabos-wissen-wer-der-babo-ist-remix-lyrics> (letzter Aufruf 24.7.2018).

⁶⁹ <https://genius.com/1338056> (letzter Aufruf 24.7.2018).

⁷⁰ Ebd.

⁷¹ Til Biermann, „Deutscher Rap und das Spiel mit dem Hass auf Juden“, in: *WELT* (2. Mai 2013), <https://www.welt.de/kultur/pop/article115793684/Deutscher-Rap-und-das-Spiel-mit-dem-Hass-auf-Juden.html>.

⁷² <https://genius.com/1311152> (letzter Aufruf 24.7.2018).

redigieren. Der alte Ersthinweis von rayo840 wurde aufgegeben. SinaTheQueen entscheidet sich für den vorgestellten Kommentar aus dem Originallied,⁷³ übernimmt Parallelstellenmethode sowie Autoritätsargument und vereinheitlicht schließlich das Wissen zur Zeile. Dass der *Duden* ‚Babo‘ 2013 zum Jugendwort kürte, wird nicht erwähnt. Der Normierungsprozess, der sich an dieser ersten Zeile zeigt, ließe sich an vielen anderen Texteinheiten belegen. Es ist darauf hinzuweisen, dass dieser nicht nur auf Kommentarebene stattfindet, sondern bereits mit der Verschriftung beginnt. Die oben vorgestellten Regeln zur Transkription überführen die sprachlich und lautlich heterogenen Lieder in die deutsche Standardschrift, die gleich einer Koine für jedermann lesbar ist.

Die Eigendynamik der digitalen Schrift verändert ebenso die Semantik der Lieder. Das kann nur beispielhaft vorgestellt werden. Nach dem Hook eröffnet der Rapper Milonair (Milad Mirza Nejad) den Remix des Freundeskreises. Zwischen dem Refrain und der ersten Strophe ist in der Hörversion ein Sample eingefügt, das im Hintergrund zwischen beiden Lied-Einheiten den Übergang schafft. In der mündlichen Performanz besitzt es eher dekorativen Charakter. Der Editor (Electrofensterheber, 67.352 IQ) klärt auf, dass es sich um das serbische Volkslied *Zumbalica* [Дзумбалуца] handelt: „Dieses Lied wurde bereits von vielen verschiedenen Personen gesungen, daher ist die Zuordnung des Samples zu einem bestimmten Künstler schwierig.“⁷⁴ Der Gattungsbegriff des Volksliedes bleibt unerwähnt. Aber die Kommentatoren betten dafür eine Hörversion des Liedes ein: „Reče čiča da me ženi/Danas hoće sutra neće“⁷⁵. In der zweidimensionalen Schrift tritt hervor, was im gesungenen Lied im Hintergrund geblieben war. Auch wird ein möglicher Sinn durch die Einspielung des serbischen Volksliedes deutlich, sobald man den ersten Vers betrachtet. Der Anspruch des Rappers, für seine Gemeinschaft zu sprechen, wird abgeleitet aus der Stimme des Volkes, welche das serbische Volkslied (bzw. das Lied, das bereits von vielen gesungen wurde) hörbar macht: „Milonair, ich verleih‘ dem Ghetto ‘ne Stimme“.⁷⁶ Das wurde 2013 durch Geisterschreiber (46.114 IQ) so erläutert: „Durch Milonair haben auch die Dealer, Unterdrückten und Armen [sic] Leute aus den Ghettos die Möglichkeit ihre Probleme und Wünsche einer breiten Masse mitzuteilen, die ihnen sonst meist verwehrt bleibt.“⁷⁷ SinaTheQueen ergänzt den zugrundeliegenden Gemeinplatz (‚Jemandem eine Stimme geben‘) und verbessert Orthographie und Zeichensetzung.

⁷³ Ebd.

⁷⁴ <https://genius.com/6707206> (letzter Aufruf 24.7.2018).

⁷⁵ Ebd.

⁷⁶ Im Kommentar heißt es dazu: „Durch Milonair haben auch die Dealer, Unterdrückten und armen Leute aus den Ghettos die Möglichkeit, ihre Probleme und Wünsche einer breiten Masse mitzuteilen, die ihnen sonst meist verwehrt bleibt.“ (<https://genius.com/1553832> [letzter Aufruf 24.7.2018]).

⁷⁷ Ebd.

Der Vorgang ist nicht ohne Ironie weil der Sänger des Volkes bzw. seiner Fangemeinde durch eben diese Gemeinde ausgelegt wird. Die traditionelle Idee des Volksliedes ist die eines kollektiv verfassten Liedes, in der sich das Volk selbst ausspricht. Auf der Seite seiner Tradierung gesellen sich neben das Volk die Philologen des Volksliedes, die sich als Agenten der Gemeinschaft verstehen. Ihr Auftrag ist es, dem gesungenen Volkslied in der Schrift eine Stimme zu verleihen. Die individuelle Poesie der Künstler kann das Moment der Agentur ebenfalls enthalten. Jeder Künstler ist dann in dem Maße Philologe, in dem er sich als Sprachrohr einer Gemeinschaft versteht bzw. konvergieren Künstler und Philologen in dem Wunsch, für andere rhetorisch und poetisch zu handeln. Auch der Rapper verleiht seiner Gemeinschaft eine Stimme. Was aber, wenn diese Gemeinschaft selbst wieder Agenten bereitstellt, die genau diesen Prozess anfangen zu reflektieren?

In der digitalen Praxis auf *Genius* wird die kollektive Verschriftung evident. Herders und Grimms Konzept einer Stimme des Volkes als eines kollektiven Überlieferungsträgers, ob nun im Lied oder im Märchen, ist hier greifbar geworden. Die digitale Verschriftung sowie die textkritische und hermeneutische Kommentierung sind tatsächliche populäre Akte und nicht mehr solche von privilegierten Philologen, die im Namen ihres Ideal-Volkes sprechen. Gemeinschaftlich getätigte digitale Transkriptions- und Kommentierungsakte begleiten die gesungenen oder gesprochenen Lieder, wodurch sich ein Kontext bildet, der diese Lieder auch als Texte lesbar und interpretierbar macht.

Selbst wenn ihre Schöpfer schriftfeindlich eingestellt sein mögen und auch der Ursprung sowie die Geschichte der Rap-Kultur bis zum Ende des 20. Jahrhunderts schriftlos waren,⁷⁸ erzwingt die Gemeinschaft der Rap-Amateure die Schriftbegleitung dieser kulturellen Formation. Der Zwang entsteht nicht aus einem kulturpolitischen Anspruch heraus, sondern aus der schieren Praxis einer nicht zu stoppenden Korpus- und Wissensproduktion. Durch die schriftliche Begleitung erst, die eine regelrechte Wissensakkumulationsmaschinerie ist, wird eine neuartige populäre Tradition gestiftet. Ab 1800 wurde das Volk zunehmend Teil der Leserschaft und veränderte den Erwartungshorizont und damit die Literatur. Seit 2000 aber erobert es die Räume der schriftbasierten Überlieferung und richtet sie nach seinen Vorstellungen neu ein.

4 Das kollektive Korpus in der Forschung. Ein Ausblick

Auf der Datenbank *Genius* reichern wirkliche Nutzer ein umfangreiches Wissen über Lieder und ihre Kontexte, ihre Medien, ihre Kultur und ihr Publikum an. Dieser Vorgang, der maßgeblich durch Verschriftung und Kommentierung entsteht, konnte hier

⁷⁸Zur Textentstehung zwischen oralen und literalen Techniken Wolbring (Anm. 1), 201–224. Heide (Anm. 1), Kap. 14, orientiert sich genauso wie Wolbring (ebd.), 219–224, und Gruber (Anm. 1), 76, an Walter J. Ong, *Oralität und Literalität. Die Technologisierung des Wortes*, Opladen 1987 [engl. 1982]. Gruber begreift mit Ong Rap als Form sekundärer Oralität (Anm. 1), 136.

nur exemplarisch vorgestellt werden. Gezeigt werden sollten in Ansätzen die auch derzeit an anderen Orten zu beobachtenden Prozesse der Selbstprofessionalisierung. Wenn philologische oder historische Schulung für die Akteure der neuen Laienkritik nicht vorausgesetzt werden kann,⁷⁹ dann ist die Einübung in wissenschaftliches, d. h. redliches Arbeiten deshalb aber nicht ausgeschlossen. Auch die Wissenskommunikation gegenwärtiger Laienphilologen differenziert institutionelle Formen und Hierarchien aus. Wie in der *Wikipedia*⁸⁰ variieren die Niveaus je nach Gegenstand und Sprachraum. Die Möglichkeit einer Professionalisierung ist auf *Genius* angelegt, indem hier textkritische Befunde oder hermeneutische und ästhetische Urteile veröffentlicht werden, um sich wieder der Kritik auszusetzen.⁸¹

Die Ausweitung des Populären vom Bereich der Aufnahme (Rezeption) auf den der Bereitstellung und Beurteilung von Inhalten hat Auswirkungen auf die künftige Überlieferung und Kanonisierung von Literatur. Insofern das populäre Lied ein der neuen populären Philologie adäquater Gegenstand ist, konkurrieren die populären Philologen mit den Akteuren der akademischen Literaturwissenschaft. Aktuell stellt sich deshalb nicht die Frage, ob die Laienpraxis epistemische Relevanz gewinnt (denn das tut sie bereits), sondern wie sie in eine Institution überführt werden kann. Die Einbindung von Laien in professionelle textkritische

⁷⁹Für den hier vorgestellten Fall von *Genius* muss diese Aussage differenziert werden. Die drei Begründer der Wissensdatenbank Tom Lehman (Ethics, Politics and Economics), Mahbod Moghadam (*Stanford Law School*) und Ilan Zechory (Religious Studies) haben in Yale und Stanford studiert, vgl. Jacob Potash, „Genius or whatever“, in: *Yale Daily News* (6. März 2015), <http://yaledailynews.com/blog/2015/03/06/geniuses-or-whatever-tom-lehman-and-ilan-zechory/> (letzter Aufruf 24.7.2018). Der Begründer des deutschen Ablegers, Tobias Wilinski, hat einen Master in Afrikawissenschaften (*HU Berlin*). Und es steht außer Zweifel, dass viele Nutzer der Plattform aus dem akademischen Raum kommen. Moghadam betont sogar die Nähe zur Literaturwissenschaft in Yale: „[...] at Yale they teach you how to do close reading. The person who is worshipped on campus is Harold Bloom, who is the father of close reading, so that’s why we came up with the integral feature. All the comments are close reading to the text and this is what makes this the ‚most Yaled-out website‘ out there“, vgl. Mahbod Moghadam, „Interview mit Yale Daily News“, in: *Yale Daily News* (7. Februar 2014), <http://yaledailynews.com/blog/2014/02/07/mahbod-moghadam-yales-rap-genius/> (24.7.2018).

⁸⁰Ehrlicher bewertet die Diskussion des spanischen *Wikipedia*-Artikels über den *Lazarillo del Tormes* (1. Oktober 2008) als „von Laienphilologen hergestellte Netzöffentlichkeit“, die „durchaus den Standards der von Habermas idealtypisch konstruierten vernunftbasierten Struktur der bürgerlichen Öffentlichkeit“ entspreche. Vgl. Hanno Ehrlicher, „Das aufgegebene Anonymat. Kritische Anmerkungen zu einer philologischen Kanonrevison aus aktuellem Anlass“, in: *Philologie im Netz* 46 (2008), 1–13, hier: 12, Anm. 4.

⁸¹Steinfeld, zwischen dem Philologen und dem Sammler trennend, vermisst bei Letzterem „die Linie der Kritik“ (Thomas Steinfeld, *Der leidenschaftliche Buchhalter. Philologie als Lebensform*, München 2004, 73). Lepper, der zwischen Philologen und Kenner trennt, ergänzt Steinfelds Professionalisierungsmerkmal der Kritik um Fachkommunikation und Institutionalisierbarkeit: „Sucht der forschende Philologe die Nähe der objektivierenden Institutionen, der Universitäten, Bibliotheken und Archive, so baut sich der Kenner seine eigene Welt.“ Kennerschaft werde identisch mit Laienphilologie. Vgl. Marcel Lepper, *Philologie zur Einführung*, Hamburg 2012, 143.

oder lexikographische Arbeitsprozesse⁸² erfolgt derzeit beispielsweise am *Deutschen Textarchiv*, wo sie der Qualitätssicherung dient.⁸³ Sicherlich ist die Reduktion auf Hilfsarbeiten wie Korrekturlesen für ein Vorhaben, das allein der Retrodigitalisierung dient, berechtigt, jedoch wäre es auch denkbar, den kritischen Raum für die hermeneutische Praxis zu öffnen. Die Angst, der kritischen Menge mehr editorische Rechte zu geben, ist nachvollziehbar. Gleichwohl zeigt ein Projekt wie *Wikisource*,⁸⁴ dass diese Menge durchaus zielorientiert arbeiten kann.

John Meier hat das Volkslied als das Lied definiert, das „volkläufig geworden“ sei.⁸⁵ Damit sollte zum einen die müßige Frage, wie das Volk als kreative Instanz zu denken sei, umgangen und zum anderen die Überlieferungsfrage auf die Mündlichkeit festgelegt werden. Die Überlieferung bezeugen nach diesem Modell Bild- und Tondokumente. Diese können auch schriftlich dokumentiert sein. Weniger aber handelt es sich bei den Zeugnissen um kritische Kommunikation. Dass das Volk seine eigenen Lieder ediert und kommentiert – daran hat Meier nicht gedacht. Der Begründer des in Freiburg angesiedelten *Deutschen Volksliedarchivs* (1914–2014, seit 2014 *Zentrum für Populäre Kultur und Musik*) hatte ein ausgeprägtes Bewusstsein dafür, dass das Volkslied die Summe seiner Überlieferungszeugnisse ist. Aus diesem Gedanken heraus ist eine wertvolle Sammlung entstanden, die ihrer digitalen Transformation harrt. Kritische Kommunikation – sowohl textkritische als auch hermeneutische – ist in der Gegenwart ein Wesensmerkmal des populären Liedes und seiner Überlieferungsdynamik geworden. Technisch forciert wird sie von der interaktiven Qualität des neuen Schriftmediums.

⁸² Patrick Sahle, *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, 3 Bde., Bd. 2: *Befunde, Theorie und Methodik*, Norderstedt 2013, 234 f. u. 259–272. Vgl. Andrea Abel/Christian M. Meyer, „Nutzerbeteiligung“, in: Annette Klosa/Carolin Müller-Spitzer (Hg.), *Internetlexikografie. Ein Kompendium*, Berlin 2016, 249–290.

Eine Pionierstudie zu diesem Thema: Raymond G. Siemens, „Shakespearean Apparatus? Explicit Textual Structures and the Implicit Navigation of Accumulated Knowledge“, in: *Text. An Interdisciplinary Annual of Textual Studies* 14 (2002, zuerst 1999), 209–240. Neuerdings: Thomas Ernst, „Vom Urheber zur Crowd, vom Werk zur Version, vom Schutz zur Öffnung? Kollaboratives Schreiben und Bewerten in den Digital Humanities“, in: Constanze Baum/Thomas Stäcker (Hg.), *Grenzen und Möglichkeiten der Digital Humanities 2015 (= Zeitschrift für digitale Geisteswissenschaften, https://doi.org/10.17175/sb001_021;* Anne Baillot, „Was tun mit der Weisheit der Massen? Moderne Philologie im digitalen Zeitalter“, in: Stefanie Stockhorst/Marcel Lepper/Vinzenz Hoppe (Hg.), *Symphilologie. Formen der Kooperation in den Geisteswissenschaften*, Göttingen 2016, 261–280.

⁸³ Die kollaborative *Qualitätssicherung des Deutschen Textarchivs (DTAQ)* gibt am 2. August 2017 folgende Zahlen an (<http://www.deustextarchiv.de/dtaq/about> [letzter Aufruf 24.7.2018]): 3.359 Werke, 69.9611 Textseiten, 1.278 Benutzer, 76.944 angelegte Tickets (68.187 behoben), 48.512 Korrektur gelesene Seiten. Im September 2016 waren es 3.037 Werke, 670.628 Textseiten, 1.080 Benutzer, 73.602 angelegte Tickets (66.925 abgeschlossen), 48.018 Korrektur gelesene Seiten.

⁸⁴ https://de.wikisource.org/wiki/Wikisource:Über_Wikisource (letzter Aufruf 24.7.2018).

⁸⁵ Meier (Anm. 48), 7.

Die Umbenennung des Deutschen Volksliedarchivs 2014 in das Freiburger *Zentrum für Populäre Kultur und Musik*⁸⁶ kann als Reaktion auf die im Digitalen durch kritische Kommunikation der Laien vergegenwärtigte populäre Musikkultur angesehen werden. In Anbetracht der populären Medienrealität wirkt das alte romantische Modell anachronistisch, und tatsächlich ließe sich heute die Privilegierung dessen, was als ‚Deutsches Volkslied‘ bezeichnet wird, innerhalb des populären Liedgutes schwer rechtfertigen.

Als ein erstes Forschungsergebnis der neuen Ausrichtung, die auch durch eine veränderte Sammelpraxis gedeckt ist, kann das hochwertige *Songlexikon. Encyclopedia of Songs. Eine Netzpublikation des Zentrums für Populäre Kultur und Musik* der *Universität Freiburg* angesehen werden. Zweihundert Songs (6/2022: 251) werden von verschiedenen Fachleuten nach Entstehung, Wirkungskontext, Form und Inhalt sowie Rezeption vorgestellt. Zugleich finden sich Angaben zu Cover-Versionen⁸⁷ sowie zur Sekundärliteratur. Für den Kommentar eines Rap-Songs wurde sogar die Wissensdatenbank *Genius* als Quelle genutzt.⁸⁸ Da es sich um bekannte Songs handelt, konkurrieren die meisten der Lieder mittlerweile mit Einträgen in der *Wikipedia*.

Zu überlegen wäre, die Sammlung und das erschlossene Wissen als Grundlage einer institutionell gesicherten und wirtschaftlich unabhängigen Wissensdatenbank zu verwenden und aus der internen Erschließungsarbeit heraus Kriterien für das kollektive Edieren und Kommentieren abzuleiten. Die wilde Praxis der Laienphilologen auf *Genius*, die auf eine studentische Initiative zurückgeht, scheint mir dafür wegweisend zu sein. Die epistemische Relevanz des Textkorpus steht fest, wie der Rückgriff der Freiburger Autoren auf die Datenbank zeigt. Aber auch in anderen Kontexten wird zunehmend auf ihr Wissen zurückgegriffen. Für die Analyse von Netzwerkstrukturen⁸⁹ und für die Untersuchung soziolinguistischer Fragen⁹⁰ hat sich die Plattform bereits als produktiv erwiesen. Da man an die API von *Genius*

⁸⁶Michael Fischer, „100 Jahre Deutsches Volksliedarchiv – Gründung des Zentrums für Populäre Kultur und Musik“, in: Michael Fischer/Tobias Widmaier (Hg.), *Lieder/Songs als Medien des Erinnerns*, Münster 2014, 9–18.

⁸⁷Um ein Beispiel zu geben, vergleiche ich die Angaben im Artikel von *Blowin' in the wind* mit denen auf *Genius*. Auf dieser Seite werden 70 Versionen verlinkt: <https://genius.com/Bob-dylan-blowin-in-the-wind-lyrics>, auf jener stehen nur sechs Titel <http://www.songlexikon.de/songlexikon/songs/blowinwind> (letzter Aufruf 24.7.2018).

⁸⁸Im Artikel zu dem Lied von Jay Z *99 Problems* steht die Quellenangabe: „<http://genius.com/Jay-z-99-problems-lyrics/> [04.01.2015]“. Vgl. Sebastian Danz, „99 Problems (Jay Z)“, in: Michael Fischer/Fernand Hörner/Christofer Jost (Hg.), *Songlexikon. Encyclopedia of Songs* (März 2017), <http://www.songlexikon.de/songs/99problems> (letzter Aufruf 24.7.2018).

⁸⁹Fields/Rhodes (Anm. 9).

⁹⁰Leonie Wiemeyer/Steffen Schaub, „Dimensions of Dissatisfaction and Dissent in Contemporary German Rap. Social Marginalization, Politics, and Identity Formation“, in: Damian J. Rivers/Andrew S. Ross (Hg.), *The Sociolinguistics of Hip-Hop as critical conscience. Dissatisfaction and Dissent*, Basingstoke 2018, 37–67.

gelangt,⁹¹ eignet sich das Netzwerk zudem dazu, große, für den Einzelnen nicht überschaubare Datenmengen auszuwerten. Ein Vergleich der nationalen Kritik- und Transkriptionskulturen auf der Plattform böte sich ebenfalls an.

Obgleich die digitalen Analysemethoden die auf Erfahrung basierende und auch bei Laien anzutreffende Kennerschaft außer Kraft gesetzt haben,⁹² sollte man den digitalen Raum nicht ausschließlich als Raum begreifen, der sich individueller Erfahrung verschlüsse. Zu dieser Sicht hat eine Fokussierung beigetragen, die unter *Digitaler Literaturwissenschaft* primär die Einbindung der Technik in die Analyse versteht. Kollektive Kritik wird jedoch weiterhin von Individuen und nicht von Maschinen getätigt, weshalb die mit ihr verbundene philologische Praxis eine subjektive Praxis bleibt. Nur nimmt sie unter den Gesetzen digitaler Medialität andere Formen an als unter den Gesetzen analoger Medialität. Die Funktionsweise des digitalen Mediums und der Unterschied zur analogen Medialität wären durch Beobachtung der laienphilologischen Praxis zu reflektieren, bevor man textkritisch, editorisch und hermeneutisch aktiv wird. So wie man mittlerweile auf der Gegenstandsebene den Begriff des ‚Volksliedes‘ durch den des ‚populären Liedes‘ ersetzt hat, wäre auch die mit dem Volkslied entstandene Methodik zu überprüfen, die auf die Präsentationstechniken der Printkultur abgestimmt war.

Ergänzung 2021

Der Nachtrag⁹³ gibt Anlass zu fünf Differenzierungen und einigen wenigen Ergänzungen von Forschungsliteratur. *Erstens* möchte ich die von mir hervor gehobene Teilhabe als Eigenschaft digitaler Schriftlichkeit dahingehend relativieren, dass diese von den infrastrukturellen und materiellen Bedingungen der Nutzer abhängt, die wiederum global stark variieren. Zudem handelt es sich um eine potenzielle Teilhabe, da im digitalen Raum eine Asymmetrie zwischen aktiven Nutzern und passiven Nutzern im Sinne traditioneller Rezipienten weiterhin bestehen bleibt und nur ein Bruchteil der zahlreichen Nutzer von *Genius* an der Plattform mitwirkt. *Zweitens* ist hinsichtlich der Dynamik der digitalen Schriftlichkeit bzw. ihrer Versatilität zu ergänzen, dass diese selten durch die Nutzer bewerkstelligt wird und dass auch die Anpassung an Endgeräte Standards gehorcht, die von den meist kommerziellen Softwareunternehmen gesetzt worden sind. Die Souveränität über die digitale Schriftlichkeit ist aufgrund fehlender informatischer Kompetenzen vieler Nutzer oder verwehrter Eingriffsmöglichkeiten in den Programmcode eher simuliert als real. *Drittens* möchte ich mit der Hybridität ein Moment digitaler Schriftlichkeit erwähnen, das in meiner Darstellung nur indirekt angesprochen ist. Diskutiert wird es beispielsweise von

⁹¹ <https://docs.genius.com> (letzter Aufruf 24.7.2018).

⁹² Vgl. Lepper (Anm. 81), 140–143.

⁹³ Für die kritische Diskussion des Aufsatzes, deren Ergebnisse in diesen Nachtrag miteingeflossen sind, danke ich Martin Bartelmus und Ann-Marie Riesner.

Henning Lobin⁹⁴, der es von dem Prinzip der Automatisierung ableitet, d. h. der Auslagerung von Rechenprozessen an Maschinen. Dadurch ist die digitale Position immer schon eine hybride Kombination menschlicher und maschineller Agentialität. Die Auswirkungen dieses Merkmals für die Analyse der Verschriftung auf *Genius* können hier nicht mehr nachgezeichnet werden. Ergänzt sei nur, dass die Agentialität solcher Plattformen immer zweigleisig ist, insofern das, was die Nutzer generieren, einer anderen Logik gehorcht als der algorithmisch erzeugte Handlungsraum, in dem sich eben diese Nutzer frei zu bewegen glauben. *Viertens* ist für den Vergleich der kollaborativen digitalen und der analogen philologischen Praxis von Bedeutung, dass nur die letztgenannte Rechtfertigungsdiskurse hervorbringt. Analoge Editoren und Herausgeber haben vornehmlich das Privileg, eine schöpferische und individuelle Autorschaft zu entwickeln. *Fünftens* weist die Interaktion zwischen den musikalischen Urhebern einerseits und den sie bzw. ihre Texte beobachtenden Nutzern andererseits Parallelen auf zur Kommunikation von literarischen Autorinnen und Autoren mit der sie beobachtenden Kritik. Auf den einschlägigen Plattformen wie Twitter interagieren sie mit ihrem Publikum; vor allem arbeiten sie gemeinsam mit der Gegenwartsliteraturwissenschaft am Verständnis ihrer Texte – eine Praxis freilich, deren Entwicklung in der Klassischen Moderne begann.

Einige Titel aus der Forschungsliteratur seien abschließend ergänzt: Meine Behauptung, die „Träger der populären Kultur“ machten durch ihren laienphilologischen Umgang mit Songs, insbesondere Rap-Songs, diese mit dem „literaturwissenschaftlichen Gattungsbegriff der ‚Lyrik‘ kompatibel“, bestätigt der Aufsatz von Philipp Böttcher⁹⁵, der auf den Seiten 78 bis 81 eine kurze Beschreibung der Plattform *Genius* gibt. – Die dezentralisierte und kollaborative Sammelpraxis erinnert, da im selben Bereich der kommerziellen Musik stattfindend, an die Praxis der frühen Internet-User um 2000, sich eine eigene Plattform für das Teilen von Musik zu schaffen. Über diese Vorgänge informiert Mercedes Bunz.⁹⁶ – Den Umstand, dass sich die „Emphase auf die Stimme [...] kompensatorisch aus dem Fehlen phonographischer Techniken erklären“ lässt, reflektiert Steffen Wallach.⁹⁷ Der Fiktion der Mündlichkeit bei den Grimms und den inhärenten Gemein-

⁹⁴Henning Lobin, *Engelbarts Traum. Wie der Computer uns Lesen und Schreiben abnimmt*, Frankfurt a.M. 2014, 77–97, 104–113, 131–141.

⁹⁵Philipp Böttcher, „Gelesener Gesang. Lyrics im Zeichen des Medienwandels“, in: Steffen Martus/Carlos Spoerhase (Hg.), *Gelesene Literatur. Populäre Lektüre im Zeichen des Medienwandels*, München 2018, 73–82.

⁹⁶Mercedes Bunz, *Vom Speicher zum Verteiler. Die Geschichte des Internet*, Berlin 2008, 18–20.

⁹⁷Steffen Wallach, „Vor der Phonographie – Herders, Arnims und Brentanos volkspoetische Gramma-Phonie“, in: Nathalie Binczek/Cornelia Epping-Jäger (Hg.), *Das Hörbuch. Praktiken audioliteralen Schreibens und Verstehens*, Paderborn 2014, 31–46.

schaftskonzepten der Märchenphilologie widmet sich Mark-Georg Dehrmann.⁹⁸ Ergänzen möchte ich zu meinen Herder-Ausführungen, dass Herder durchaus auch individuell verfasste Lieder als volksmäßig verstanden hat und, bedingt durch fehlende Zeugnisse, kaum den globalen Süden für sein global angelegtes Volksliedprojekt berücksichtigen konnte. Im Aufsatz zu Herder und Gräter⁹⁹ diskutiere ich Herders Theorie und Praxis der Lied-Überlieferung im historischen Kontext. – Zur Frage der Veränderbarkeit von digitaler Schrift sei nachträglich auf das klassische Buch von N. Katherine Hayles verwiesen, die digitale Texte als „text-as-flickering-image“ bezeichnet.¹⁰⁰ – Zusätzlich zu den genannten Arbeiten zur Partizipationskultur, Interaktivität und kollektiver Korpus-Arbeit verweise ich auf die Studien der Theoriebegründer Howard Rheingold¹⁰¹ sowie Ramón Reichert.¹⁰² Die gemeinschaftliche Textumgangsform ist wiederum Teil einer Kultur der Digitalität, wie sie von Felix Stalder¹⁰³ beschrieben wird. Neben Referentialität und Algorithmizität spricht er von Gemeinschaftlichkeit als einem Konzept, mit dem die Neuordnung von Technik und Gesellschaft erfasst werden kann. Gemeint ist ein kollektiver „Referenzrahmen“, durch den „Bedeutungen stabilisiert, Handlungsoptionen generiert und Ressourcen zugänglich gemacht werden.“¹⁰⁴ Zugleich werde „soziales Handeln [...] in zunehmend komplexere Technologien eingebettet, ohne die diese Prozesse kaum zu denken und schon gar nicht zu bewerkstelligen wären.“¹⁰⁵ Arbeiten zum kollaborativen Schreiben sind in den letzten Jahren zahlreich erschienen, allerdings gibt es kaum etwas zur kollaborativen philologischen Praxis. Am ehesten wäre noch die Wikipedia-Forschung zu Rate zu ziehen.¹⁰⁶

⁹⁸ Mark-Georg Dehrmann, „Die Austreibung der Schrift durch die Schrift. Zur philologisch-historischen Reflexion von Mündlichkeit nach 1800 am Beispiel der Grimmschen Kinder- und Hausmärchen“, in: *Fabula* 55/1 (2014), 153–170.

⁹⁹ Alexander Nebrig, „Das Lied und die Schrift. Überlieferung bei Johann Gottfried Herder und Friedrich David Gräter“, in: *Jahrbuch des Österreichischen Volksliedwerkes* 69 (2020), 17–31.

¹⁰⁰ N. Katherine Hayles, *How We Became Posthuman. Virtual Bodies in Cybernetics, Literatur and Informatics*, Chicago 1999, Zitat S. 26.

¹⁰¹ Howard Rheingold, „Smart Mobs. Die Macht der mobilen Vielen“ [2002], in: Karin Bruns/Ramón Reichert (Hg.), *Neue Medien*, Bielefeld 2007, 359–370.

¹⁰² Ramón Reichert, *Die Macht der Vielen. Über den neuen Kult der digitalen Vernetzung*, Bielefeld 2013.

¹⁰³ Felix Stalder, *Kultur der Digitalität*, Frankfurt a.M. 2016.

¹⁰⁴ Ebd., 13.

¹⁰⁵ Ebd., 11.

¹⁰⁶ Kerstin Kallass, *Schreiben in der Wikipedia. Prozesse und Produkte gemeinschaftlicher Textgenese*, Wiesbaden 2015.

Literatur

- Abel, Andrea/Meyer, Christian M., „Nutzerbeteiligung“, in: Annette Klosa/Carolin Müller-Spitzer (Hg.), *Internetlexikografie. Ein Kompendium*, Berlin 2016, 249–290.
- Baillot, Anne, „Was tun mit der Weisheit der Massen? Moderne Philologie im digitalen Zeitalter“, in: Stefanie Stockhorst/Marcel Lepper/Vinzenz Hoppe (Hg.), *Symphilologie. Formen der Kooperation in den Geisteswissenschaften*, Göttingen 2016, 261–280.
- Bartl, Andrea/Behmer, Markus (Hg.), *Die Rezension. Aktuelle Tendenzen der Literaturkritik*, Würzburg 2016.
- Bausinger, Hermann, „Gräters Beitrag zur Volksliedforschung“, in: *Jahrbuch Württembergisch Franken* 52 (1968), 73–94.
- Biermann, Til, „Deutscher Rap und das Spiel mit dem Hass auf Juden“, in: *WELT* (2. Mai 2013), <https://www.welt.de/kultur/pop/article115793684/Deutscher-Rap-und-das-Spiel-mit-dem-Hass-auf-Juden.html>.
- Böck, Sebastian (Hg.), *Lesen X.0. Rezeptionsprozesse in der digitalen Gegenwart*, Göttingen 2017.
- Bolter, Jay D., „Das Internet in der Geschichte der Technologien des Schreibens“, in: Sandro Zanetti (Hg.), *Schreiben als Kulturtechnik. Grundagentexte*, Frankfurt a.M. 2015, 318–337.
- Bolter, Jay D., „Digitale Schrift“, in: Gernot Grube/Werner Kogge/Sybille Krämer, *Schrift. Kulturtechnik zwischen Auge, Hand und Maschine*, München 2005, 453–467.
- Borchmeyer, Dieter, „Poesie für das Ohr – Tönende versus gelesene Dichtung. Zur Geschichte eines Strukturproblems von Klopstock bis Wagner“, in: Walter Pape (Hg.), *Das „Wunderhorn“ und die Heidelberger Romantik. Mündlichkeit, Schriftlichkeit, Performanz*, Tübingen 2005, 207–223.
- Böttcher, Philipp, „Gelesener Gesang. Lyrics im Zeichen des Medienwandels“, in: Steffen Martus/Carlos Spoerhase (Hg.), *Gelesene Literatur. Populäre Lektüre im Zeichen des Medienwandels*, München 2018, 73–82.
- Bunz, Mercedes, *Vom Speicher zum Verteiler. Die Geschichte des Internet*, Berlin 2008, 18–20.
- Büttner, Urs, *Poesis des ‚Sozialen‘. Achim von Arnims frühe Poetik bis zur Heidelberger Romantik (1800–1808)*, Berlin 2015.
- Czollek, Maximilian, „Jubiljahre. Zur Aktualität theologischer Topoi in der deutschen Gegenwartsliteratur“, in: Dörte Linke/Florian Priesemuth/Rosa Schinagl (Hg.), *Sprachen des Unsagbaren. Kulturelle Figurationen: Artefakte, Praktiken, Fiktionen*, Wiesbaden 2017, 243–266.
- Danz, Sebastian, „99 Problems (Jay Z)“, in: Michael Fischer/Fernand Hörner/Christofer Jost (Hg.), *Songlexikon. Encyclopedia of Songs* (März 2017), <http://www.songlexikon.de/songs/99problems>.
- Dehrmann, Mark-Georg, „Die Austreibung der Schrift durch die Schrift. Zur philologisch-historischen Reflexion von Mündlichkeit nach 1800 am Beispiel der Grimmschen Kinder- und Hausmärchen“, in: *Fabula* 55/1 (2014), 153–170.
- Dommann, Monika, *Autoren und Apparate. Die Geschichte des Copyrights im Medienwandel*, Frankfurt a.M. 2014.
- Eagleton, Terry, *Einführung in die Literaturtheorie*, Stuttgart 2012 (zuerst dt. 1988).
- Ehrlicher, Hanno, „Das aufgegebenen Anonymat. Kritische Anmerkungen zu einer philologischen Kanonrevision aus aktuellem Anlass“, in: *Philologie im Netz* 46 (2008), 1–13.
- Ernst, Thomas, „User Generated Content‘ und der Leser-Autor als ‚Prosumer‘. Potenziale und Probleme der Literaturkritik in Sozialen Medien“, in: Christina Gansel/Heinrich Kaulen (Hg.), *Literaturkritik heute. Tendenzen – Traditionen – Vermittlung*, Göttingen 2015, 95–111.
- Ernst, Thomas, „Vom Urheber zur Crowd, vom Werk zur Version, vom Schutz zur Öffnung? Kollaboratives Schreiben und Bewerten in den Digital Humanities“, in: Constanze Baum/Thomas Stäcker (Hg.), *Grenzen und Möglichkeiten der Digital Humanities 2015 (= Zeitschrift für digitale Geisteswissenschaften)*, DOI https://doi.org/10.17175/sb001_021.
- Escarpit, Robert, *Die Revolution des Buches*, Gütersloh 1967.

- Fields, Ben/Rhodes, Christophe, „Listen To Me – Don’t Listen To Me. What Communities of Critics Tell Us About Music. ISMIR 2016“, in: Michael I. Mandel/Johanna Devaney/Douglas Turnbull u.a. (Hg.), *Proceedings of the 17th International Society for Music Information Retrieval (ISMIR) Conference 2016, New York City, August 7–11, 2016*, New York 2016, 199–205, https://wp.nyu.edu/ismir2016/wp-content/uploads/sites/2294/2016/07/173_Paper.pdf.
- Fischer, Michael, „100 Jahre Deutsches Volksliedarchiv – Gründung des Zentrums für Populäre Kultur und Musik“, in: Michael Fischer/Tobias Widmaier (Hg.), *Lieder/Songs als Medien des Erinnerns*, Münster 2014, 9–18.
- Fohrmann, Jürgen/Voßkamp, Wilhelm (Hg.), *Von der gelehrten zur disziplinären Gemeinschaft*, Stuttgart 1987.
- Gamper, Michael, „Kollektive Autorschaft/Kollektive Intelligenz. 1800–2000“, in: *Jahrbuch der deutschen Schillergesellschaft* 45 (2001), 380–403.
- Gray, John, *Hip-Hop Studies. An international bibliography and reference guide*, New York 2016.
- Grimm, Jacob, „Gedanken. Wie sich die Sagen zur Poesie und Geschichte verhalten“, in: *Zeitung für Einsiedler* (4./7. Juni 1808), 152–156.
- Gruber, Johannes, *Performative Lyrik und lyrische Performance. Profilbildung im deutschen Rap*, Bielefeld 2016.
- Haas, Daniel, „Chabos wissen, wer der Babo ist. Gaumenfreuden aus dem Getto. Der Rapper Haftbefehl macht Metasprache aus prekärem Slang“, in: *FAZ.NET* (31. Januar 2013, <http://www.faz.net/aktuell/feuilleton/pop/rapper-haftbefehl-chabos-wissen-wer-der-babo-ist-12046385.html>).
- Hayles, N. Katherine, *How We Became Posthuman. Virtual Bodies in Cybernetics, Literatur and Informatics*, Chicago 1999, Zitat S. 26.
- Heide, Markus, *Rap zwischen Mündlichkeit, Schriftlichkeit und Technologie. Eine Verortung anhand deutschsprachiger Raplieder*, Erlangen 2011.
- Herder, Johann G., *Journal meiner Reise im Jahr 1769*, hg. von Katharina Mommsen, Stuttgart 1976.
- Herder, Johann G., *Werke in zehn Bänden*, Bd. 3: *Volkslieder, Übertragungen, Dichtungen*, hg. von Ulrich Gaier, Frankfurt a.M. 1990.
- Hügel, Hans-Otto, „Einführung“, in: Ders. (Hg.), *Handbuch populäre Kultur. Begriffe, Theorien und Diskussionen*, Stuttgart 2003, 1–22.
- Jenkins, Henry, *Fans, Bloggers, and Gamers. Exploring Participatory Culture*, New York 2006.
- Kallass, Kerstin, *Schreiben in der Wikipedia. Prozesse und Produkte gemeinschaftlicher Textgenese*, Wiesbaden 2015.
- Kessler, Sarah, „Ben Horowitz Says \$15 Million For Rap Genius Will Create ‚A Talmud For The Internet‘“, in: *FastCompany* (10. März 2012), <http://www.fastcompany.com/3001879/ben-horowitz-says-15-million-rap-genius-will-create-talmud-internet>.
- Kristeva, Julia, „Bakhtine, le mot, le dialogue et le roman“, in: *Critique. Revue générale des publications françaises et étrangères* 23 (1967), 438–465.
- Kuhn, Axel/Hagenhoff, Svenja, „Digitale Lesemedien“, in: Ursula Rautenberg/Ute Schneider (Hg.), *Lesen. Ein interdisziplinäres Handbuch*, Berlin 2015, 361–380.
- Landow, George P., *Hypertext 3.0. Critical Theory and New Media in a Global Era*, Baltimore 2006 (1992).
- Langenbucher, Wolfgang R., „Die Demokratisierung des Lesens in der zweiten Leserevolution“, in: Herbert Göpfert/Ruth Meyer/Ludwig Muth u.a. (Hg.), *Lesen und Leben*, Frankfurt a.M. 1975, 12–35.
- Lepper, Marcel, *Philologie zur Einführung*, Hamburg 2012.
- Linder-Beroud, Waltraud, *Von der Mündlichkeit zur Schriftlichkeit? Untersuchungen zur Interdependenz von Individualdichtung und Kollektivlied*, Frankfurt a.M. 1990.
- Lo, Andrea, „How a rap lyric inspired a multimillion-dollar online platform“, in: *CNN Money* (20. Juni 2018), <http://money.cnn.com/2018/06/27/smallbusiness/genius-annotation-platform/index.html>.
- Lobin, Henning, *Engelbarts Traum. Wie der Computer uns Lesen und Schreiben abnimmt*, Frankfurt a.M. 2014, 77–97, 104–113, 131–141.

- Maina, Job King'ori/Suleman, Hussein, „Enhancing Digital Heritage Archives Using Gamified Annotations“, in: Robert B. Allen/Jane Hunter (Hg.), *Digital Libraries: Providing Quality Information*, Cham 2015, 169–179.
- Martus, Steffen, *Werkpolitik. Zur Literaturgeschichte kritischer Kommunikation vom 17. bis ins 20. Jahrhundert mit Studien zu Klopstock, Tieck, Goethe und George*, Berlin/New York 2007.
- Meier, John, „Einführung“, in: Ders. (Hg.), *Das deutsche Volkslied*, Bd. 1, Leipzig 1935, 7–34.
- Moghadam, Mahbod, „Interview mit Yale Daily News“, in: *Yale Daily News* (7. Februar 2014), <http://yaledailynews.com/blog/2014/02/07/mahbod-moghadam-yales-rap-genius/>.
- Nebrig, Alexander, „Die Welt als Lied. Der globale Anspruch von Herders Volksliedern“, in: Christian Moser/Linda Simonis (Hg.), *Figuren des Globalen. Weltbezug und Welterzeugung in Literatur, Kunst und Medien*, Göttingen 2014, 315–325.
- Nebrig, Alexander, „Das Lied und die Schrift. Überlieferung bei Johann Gottfried Herder und Friedrich David Gräter“, in: *Jahrbuch des Österreichischen Volksliedwerkes* 69 (2020), 17–31.
- Ngai, Sianne, *Our Aesthetic Categories Zany, Cute, Interesting*, Cambridge, MA, 2012.
- Ong, Walter J., *Oralität und Literalität. Die Technologisierung des Wortes*, Opladen 1987 [engl. 1982].
- Potash, Jacob, „Genius or whatever“, in: *Yale Daily News* (6. März 2015), <http://yaledailynews.com/blog/2015/03/06/geniuses-or-whatever-tom-lehman-and-ilan-zechory/>.
- Reichert, Ramón, *Die Macht der Vielen. Über den neuen Kult der digitalen Vernetzung*, Bielefeld 2013.
- Renner, Kaspar, „„Des Vaters Wille“. Nachlassbewusstsein und Werkpolitik in der Familie Herder“, in: Kai Sina/Carlos Spoerhase (Hg.), *Nachlassbewusstsein. Literatur, Archiv, Philologie 1750–2000*, Göttingen 2017, 179–216.
- Rheingold, Howard, „Smart Mobs. Die Macht der mobilen Vielen“ [2002], in: Karin Bruns/Ramón Reichert (Hg.), *Neue Medien*, Bielefeld 2007, 359–370.
- Rockenberger, Annika, „„Paratext“ und Neue Medien. Probleme und Perspektiven eines Begriffs-transfers“, in: *Philologie im Netz* 76 (2016), 20–60.
- Röllerke, Heinz, *Nebeninschriften. Brüder Grimm – Arnim und Brentano – Droste-Hülshoff. Literarhistorische Studien*, Bonn 1980, 1–15.
- Sämtliche digitalen Referenzen wurden letztmalig am 24.7.2018 eingesehen.
- Sahle, Patrick, *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, 3 Bde., Bd. 2: *Befunde, Theorie und Methodik*, Norderstedt 2013.
- Schäfer, Jörgen, „Netzliteratur“, in: Natalie Binzcek/Till Dembeck/Jörgen Schäfer (Hg.), *Medien der Literatur. Ein Handbuch*, Berlin 2013, 481–501.
- Schiffirin, André, *Verlage ohne Verleger. Über die Zukunft der Bücher*, übers. von Gerd Burger, Berlin 2000.
- Schiller, Friedrich, *Die Räuber. Ein Schauspiel*, Frankfurt a.M./Leipzig 1781.
- Siemens, Raymond G., „Shakespearean Apparatus? Explicit Textual Structures and the Implicit Navigation of Accumulated Knowledge“, in: *Text. An Interdisciplinary Annual of Textual Studies* 14 (2002, zuerst 1999), 209–240.
- Stalder, Felix, *Kultur der Digitalität*, Frankfurt a.M. 2016.
- Stanley, Lawrence A. (Hg.), *Rap. The lyrics*, New York 1992.
- Steinfeld, Thomas, *Der leidenschaftliche Buchhalter. Philologie als Lebensform*, München 2004.
- Wallach, Steffen, „Vor der Phonographie – Herders, Arnims und Brentanos volkspoetische Gramma-Phonie“, in: Nathalie Binzcek/Cornelia Epping-Jäger (Hg.), *Das Hörbuch. Praktiken audioliteralen Schreibens und Verstehens*, Paderborn 2014, 31–46.
- Wegmann, Thomas, „Warentest und Selbstmanagement. Literaturkritik im Web 2.0 als Teil nachbürgerlicher Wissens- und Beurteilungskulturen“, in: Matthias Beilein/Claudia Stockinger/Simone Winko (Hg.), *Kanon, Wertung und Vermittlung. Literatur in der Wissensgesellschaft*, Berlin 2012, 279–291.
- Wiedeman, Reeves, „Genius Idea“, in: *New York Magazine* (29. Dezember 2014, online 4. Januar 2015), <http://nymag.com/daily/intelligencer/2014/12/genius-minus-the-rap.html>.
- Wiedemann, Carolin, *Kritische Kollektivität im Netz. Anonymous, Facebook und die Kraft der Affizierung in der Kontrollgesellschaft*, Bielefeld 2016.

- Wiemeyer, Leonie/Schaub, Steffen, „Dimensions of Dissatisfaction and Dissent in Contemporary German Rap. Social Marginalization, Politics, and Identity Formation“, in: Damian J. Rivers/Andrew S. Ross (Hg.), *The Sociolinguistics of Hip-Hop as critical conscience. Dissatisfaction and Dissent*, Basingstoke 2018, 37–67.
- Wirth, Uwe, „Performative Rahmung, parergonale Indexikalität. Verknüpfendes Schreiben zwischen Herausgeberschaft und Hypertextualität“, in: Ders. (Hg.), *Performanz. Von der Sprachphilosophie zu den Kulturwissenschaften*, Frankfurt a.M. 2002, 403–433.
- Wolbring, Fabian, *Die Poetik des deutschsprachigen Rap*, Göttingen 2015.
- Wolf, Friedrich A., *Prolegomena zu Homer* [lat. 1795], übers. von Hermann Muchau, Leipzig 1908 (lat. 1795).
- Wyss, Ulrich, *Die Wilde Philologie. Jacob Grimm und der Historismus*, München 1979.

Online-Ressourcen

Genius, <https://genius.com>.

Qualitätssicherung des Deutschen Textarchivs (DTAQ), <http://www.deustextarchiv.de/dtaq>.

Songtexte, <https://www.songtexte.com>.

Wikipedia. *Die freie Enzyklopädie*, „Lazarillo del Tormes“ (1. Oktober 2008), https://de.wikipedia.org/w/index.php?title=Spezial:Zitierhilfe&page=Lazarillo_de_Tormes&id=186916228.

Wikisource, <https://de.wikisource.org/wiki/Hauptseite>.

YouTube, <https://www.youtube.com>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Tausend Tode, tausend Autoren, tausend Texte? Zur Textualität digitaler Literatur

Julia Nantke

1 Einleitung, Verortung und Zielsetzung

Die Digitalisierung der Textproduktion, -distribution und -rezeption im Rahmen des aktuellen Medienwandels übt, das ist unbestritten, einen kaum zu überschätzenden Einfluss auf den wissenschaftlichen und privaten Umgang mit Texten sowie auf unsere Text-basierte Kultur als Ganze aus. Die davon betroffenen Bereiche interagieren auf vielfältige Weise und haben so an einer Remodellierung zentraler geisteswissenschaftlicher Parameter teil.

Digitale Literatur – so die Ausgangsthese – stellt einen besonders geeigneten Ausgangspunkt für eine Betrachtung der Auswirkungen der Digitalisierung auf die zentralen, das literarische und literaturwissenschaftliche Feld organisierenden Parameter dar: Sie ist selbst Abbild und Reflexionsmedium der Veränderungen, indem sie sich produktions- wie rezeptionsästhetisch in einem Spannungsfeld von Traditionslinien der analogen Literatur und einer angesichts von digitalen Kommunikationskanälen und Formaten „neue[n], spezifische[n] kulturelle[n] Umwelt“¹ situiert, die mittlerweile durch die ubiquitäre Benutzung von Computern und Internet in quasi allen Lebensbereichen ebenfalls einen spezifischen medialen Erwartungshorizont ausgeprägt hat. In literaturwissenschaftlicher Perspektive bildet digitale Literatur den zentralen Schnittpunkt sämtlicher von den Phänomenen der Digitalisierung erfasster Bereiche von der Produktion und Publikation bis hin zur Rezeption, Analyse und Archivierung.

¹ Felix Stalder, *Kultur der Digitalität*, Frankfurt a. M. 2016, 95.

J. Nantke (✉)
Universität Hamburg, Hamburg, Deutschland
E-Mail: julia.nantke@uni-hamburg.de

Ausgehend von diesen Überlegungen wird im Folgenden anhand zweier aktueller Beispiele der digitalen Literatur das Wechselverhältnis von Literaturproduktion, neuen medialen und technischen Möglichkeiten der Produktion und Rezeption sowie diskursiver Verortung in Bezug auf die heuristische Konzeptualisierung des Textes betrachtet. Der Textbegriff erweist sich bereits hinsichtlich seiner Definition und Verortung in analoger Literatur als äußerst komplex und ist spätestens seit den (post-)strukturalistischen Interventionen der 1960er Jahre vielfältigen Zuschreibungen unterworfen. Eine umfassende Rekonstruktion der verschiedenen Positionen kann daher nicht das Ziel dieses Beitrags sein. Vielmehr sollen anhand der spezifischen an den zugrunde gelegten Beispielen beobachteten Phänomene (Un-)Vereinbarkeiten und Verschiebungen herausgestellt werden, die sich in Bezug auf ‚Text‘ und ‚Textualität‘ als Rezeptions-, Editions- und Analysekonzepte der Literaturwissenschaft feststellen lassen. Bei der Verortung derselben spielen sowohl das Verhältnis zu anknüpfenden Konzepten als auch Differenzierungen verschiedener Ebenen innerhalb des Konzepts eine zentrale Rolle. Deshalb wird ‚Text‘ nicht als autarke Größe begriffen, sondern als Schnittpunkt eines interdependenten heuristischen Feldes betrachtet, welches sich in der Relationierung verschiedener, mit dem ‚eigentlichen‘ Text verknüpfter Instanzen (Autor, Rezipient, Herausgeber) und Parameter (etwa Medium, Werk, Paratext, Fassung) konstituiert, die an der Formierung des Konzepts ‚Text‘ teilhaben. Der Beitrag fokussiert daher insbesondere die Verschiebungen hinsichtlich dieser Instanzen und Parameter, um auf diese Weise übergeordnete Aussagen in Bezug auf die Textualität digitaler Literatur treffen zu können.

Die folgenden Überlegungen schließen ausgehend von der Ebene der konkreten Beispiele aufgrund der beschriebenen Wechselwirkungen produktions- und rezeptionsästhetische Perspektivierungen ein. Auch wenn Bajohr zu Recht betont, dass aus Rezipientenperspektive nicht zwangsläufig zwischen genuin digitaler und digitalisierter Literatur zu unterscheiden ist,² wirken sich die Parameter einer rein elektronischen Publikation in entscheidender Weise auf die rezipientenseitige Wahrnehmung und den Umgang mit Literatur aus. Eine einseitige Fokussierung der Produzentenperspektive muss daher unterkomplex in Bezug auf die anvisierte Fragestellung bleiben.

Dennoch sind die literarischen Beispiele eingangs hinsichtlich ihres Verhältnisses zur ‚Digitalität‘ zu differenzieren, da diese Verortungen Einfluss auf die jeweilige Perspektivierung nehmen. Beide Beispiele stellen genuin digitale Literatur in dem Sinne dar, dass sie explizit für die elektronische Publikation produziert und ausschließlich digital veröffentlicht wurden. Sie sind also produktions- wie rezeptionsästhetisch entscheidend von den Bedingungen digitaler Medialität geprägt. Während der Vertrieb des im Frohmann-Verlag

²Vgl. Hannes Bajohr, „Das Reskilling der Literatur. Einleitung zu Code und Konzept“, in: Ders. (Hg.), *Code und Konzept. Literatur und das Digitale*, Berlin 2016, 9.

erschienenen *Tausend Tode schreiben* allerdings im E-Book-Format erfolgt³ und die literarischen Inhalte der Publikation im Einzelnen nicht erkennen lassen, mit welchen (analogen oder digitalen) Mitteln sie erzeugt wurden, stellt sich die Situation für die Arbeiten des Künstlerkollektivs *Ox0a* anders dar:⁴ Diese erscheinen auf der gleichnamigen Webseite des Kollektivs und sind nicht nur für die elektronische Publikation, sondern maßgeblich mit digitalen Mitteln, nämlich mithilfe von Algorithmen, teilautomatisch generiert. Diese Unterschiede deuten die Heterogenität der (hier) unter dem Schirm der ‚digitalen Literatur‘ gefassten Gegenstände an.⁵ Sie bilden deshalb entscheidende Ansatzpunkte für eine Sensibilisierung der folgenden Überlegungen in Bezug auf die mit der Digitalisierung einhergehende Diversifizierung von Produktions- und Rezeptionsbedingungen sowie für die Suche nach strukturellen Konstanten. Gleichzeitig verweisen sie auf die Begrenztheit dieser Untersuchung, deren Beispiele nicht annähernd das gesamte Feld der digitalen Literatur abdecken (können). Der maßgeblichen Fokussierung der folgenden Analyse auf zwei Beispiele aus dem umfangreichen und vielgestaltigen Feld digitaler Literatur(en) liegt die Annahme zugrunde, dass neue literarische Formate eine genaue Beobachtung ihrer Kompositionsformen und ‚Funktionsweisen‘ sowie ihrer medialen, technischen und kulturellen Bedingungen bedürfen, da sich nur auf diese Weise differenzierte Aussagen über Veränderungen der Literatur und des sie umgebenden Systems treffen lassen.

2 Dynamische Synthesen: Digitale Literatur zwischen wuchernden Einzeltexten und polyphonem Gesamtkorpus

Die Korpora von *Tausend Tode* und *Ox0a* zeichnen sich durch Gemeinsamkeiten aus, welche sie grundsätzlich von analogen Mediatisierungen abgrenzen und so selbstreferenziell auf die medienspezifische „(post-)digitale Situation“⁶ der Texte verweisen. Zum einen sprengt ihr bloßer Umfang die herkömmlichen Ausmaße einer Buchpublikation. Dies gilt für *Ox0a* in verschärftem Maße, wenn man den übergreifenden paratextuellen Rahmen der Webseite berücksichtigt, deren spezifische Struktur sich auch jenseits des Textumfangs kaum analog darstellen ließe.

Zum anderen stellt die bisherige Bezeichnung von *Tausend Tode* und *Ox0a* als zwei Beispiele für digitale Literatur gewissermaßen eine Verlegenheitslösung dar, die gleich zu Beginn auf den Kern der Herausforderung des Textbegriffs durch die digitale Literatur zielt: Weder hinsichtlich *Tausend Tode* noch in

³ Christiane Frohmann (Hg.), *Tausend Tode schreiben*, Version 3.1., Berlin 2015.

⁴ <http://Ox0a.li/de/> (letzter Aufruf 4.5.2018).

⁵ Zu alternativen Benennungen und deren Differenzierung vgl. Heiko Zimmermann, *Autorschaft und digitale Literatur. Geschichte, Medienpraxis und Theoriebildung*, Trier 2015, 9–14.

⁶ Bajohr (Anm. 2), 9.

Bezug auf *Ox0a* kann ohne Weiteres von jeweils einem einzigen Text gesprochen werden. Umgekehrt weisen beide Beispiele in unterschiedlicher Form und in verschiedenem Maße kompositorische Momente auf, welche eine Wahrnehmung als Einheit nahelegen. Parameter wie Autorschaft, Titel und paratextuelle Einfassung, die bei analoger Literatur gemeinhin die Textgrenze markieren, bewirken bei den hier verhandelten Beispielen ein Oszillieren zwischen Einzeltexten und Gesamttext. Diese Gemeinsamkeiten stehen in Wechselwirkung mit Verschiebungen hinsichtlich der Gestaltung von Zusammenhängen mittels Literatur, welche teilweise aus veränderten Produktions- und Publikationsbedingungen hervorgehen, aber auch auf eine gewandelte kulturelle Praxis verweisen und sich auf die Entwicklung adäquater Rezeptionsstrategien auswirken.

2.1 Referenzialität als synthetisierendes Moment

Die schon im Hinblick auf analoge (post-)moderne Literatur äußerst fragwürdige Orientierung an klassischen Gattungseinteilungen und an der Differenzierung in fiktionale und faktuale Texte erweist sich aus verschiedenen Gründen für *Ox0a* und *Tausend Tode* als endgültig obsolet. Hinzu kommt für beide Korpora eine Fraktionierung der Instanz des Autors, welche die einheitsstiftende Funktion von Autorschaft in Bezug auf den Text auf unterschiedliche Weise schwächt. Dem Verlust dieser der Lektüre und Analyse von Literatur Orientierung bietenden Referenzfolien stehen in beiden Beispielen synthetisierende Bewegungen in Form struktureller und inhaltlicher Bezugspunkte entgegen.

Die Publikation *Tausend Tode*, 2014 von Christiane Frohmann in ihrem gleichnamigen Verlag begonnen, hat es sich zum Ziel gesetzt, „in Form von tausend kurzen Texten tausend höchst subjektive Ansichten auf den Tod zu versammeln, damit diese zusammenwirkend einen transpersonalen Metatext über den Tod schreiben“.⁷ Diese Texte werden sukzessive in Form von sich aktualisierenden Versionen der Sammlung veröffentlicht, wobei der Käufer einer Version ein Abonnement auf alle weiteren publizierten Versionen erwirbt. In der aktuellen Version 3.1 beläuft sich die Zahl der Beiträge – alle Texte stammen (bisher) von unterschiedlichen Autoren – auf 425, wobei anhand des Inhaltsverzeichnisses abzulesen ist, welche neuen Texte in jeder neuen Version jeweils hinzugekommen sind. Angestrebt sind also zunächst einmal tausend einzelne Texte. Die einleitenden Worte von Frohmann verweisen allerdings bereits auf die Spannung, welche innerhalb der Publikation zwischen Einzel- und Gesamttext entsteht.⁸ Auch wenn die Texte in Inhalt, Stil und Genre sehr heterogen

⁷ So Frohmann (Anm. 3), Vorwort.

⁸ Eine ähnliche Anlage weisen die beiden digitalen ‚Erinnerungsprojekte‘ von Jan U. Hasecke, *Das Generationenprojekt* (seit 1997), <https://www.generationenprojekt.de> und Guido Grigat, *23:40* (seit 1997), <http://www.dreiundzwanzigvierzig.de> (letzter Aufruf der Links 4.5.2018) auf, deren einheitsstiftende Grundlage jeweils bestimmte Zeiträume bilden, die es mit Erinnerungen zu füllen gilt. Während das *Generationenprojekt* auf die kollektiven Erinnerungen

sind, bildet das Thema ‚Tod‘, welches (in der Regel) vor der Produktion der Texte vom Verlag gesetzt wurde, das zentrale Konzept, unter dem die Einzeltexte geschrieben wurden und werden und welches sie zur kompositorischen Einheit zusammenschließt.⁹ Der Umfang des Korpus lässt dabei den Einzeltext zugunsten des polyphonen Gesamtwerks und der Wechselwirkungen zwischen dessen einzelnen Teilen in den Hintergrund treten.

Im Gegensatz zum ‚überzeitlichen‘ Grundsatzthema von *Tausend Tode* zeichnen sich die Einzeltexte von *Ox0a* insbesondere durch ihren Bezug zu aktuellen gesellschaftlich relevanten und trivialen Themen sowie durch explizite Referenzen auf bestehende literarische Werke aus. Das gemeinsame Verfahren des *Sampling* erzeugt dabei anhand der konzeptuell gesetzten Auswahl- und Kombinationskriterien eine spezifische Ordnung. So bietet *Glaube, Liebe, Hoffnung* eine Auswahl der Facebook-Kommentare von Pegida-Anhängern, die „mit den paulinischen Tugenden aus dem Hohelied der Liebe konfrontiert werden“,¹⁰ indem die Posts via Algorithmus nach den drei im Titel genannten Termini gefiltert wurden und das Ergebnis den Text bildet. Das Korpus unter dem Titel *Chicken Infinite* setzt sich zusammen aus „[c]ooking recipes from the web [which] have been collected and mixed randomly together.“¹¹

Das Herstellen von und Einbinden in Netzwerkstrukturen sowie das Ausschneiden aus und die Remodellierung von bestehenden Netzwerken, durch die sich die Produktionspraktiken von *Tausend Tode* und *Ox0a* auszeichnen, können mit Stalder als zentrale Elemente des für die „Kultur der Digitalität“ spezifischen Paradigmas der Referenzialität angesehen werden.¹² Die aus den Produktionsverfahren resultierenden Kompositionen reagieren auf den „Flickenteppich verbundener, konkurrierender oder auch einfach voneinander getrennter Sphären

zu bedeutenden Ereignissen zielt, werden in 23:40 persönliche Erinnerungen beschrieben (vgl. Roberto Simanowski, *Interfictions. Vom Schreiben im Netz*, Frankfurt a. M. 2002, 37–43). Die beiden Projekte markieren damit die Pole ‚kollektive Erfahrung‘ und ‚radikale Subjektivität‘, zwischen denen sich *Tausend Tode* (Frohmann [Anm. 3]) konzeptuell bewegt.

⁹Natürlich lässt sich nicht mit Sicherheit sagen, ob die Texte erst im Anschluss an die Setzung des Konzepts geschrieben wurden. Es könnte sich teilweise auch um bereits zuvor Produziertes handeln. Von den aktuell 425 Texten wurden aber – wie am Ende der Publikation ausgewiesen – lediglich 18 zuvor oder gleichzeitig über andere Kanäle publiziert, sodass grundsätzlich von einer starken Dependenz von Konzept und Schreiben auszugehen ist.

¹⁰Hannes Bajohr/Gregor Weichbrodt, *Glaube Liebe Hoffnung* (2015), <http://Ox0a.li/de/text/glaube-liebe-hoffnung/> (letzter Aufruf 4.5.2018).

¹¹Gregor Weichbrodt, *Chicken Infinite* (2014), <http://Ox0a.li/de/text/chicken-infinite/> (letzter Aufruf 4.5.2018).

¹²„Eine, wenn nicht die grundlegende Methode, mit der Menschen – alleine und in Gruppen – an der kollektiven Verhandlung von Bedeutung teilnehmen, besteht in der Kultur der Digitalität darin, Bezüge herzustellen.“ Stalder (Anm. 1), 96; vgl. insgesamt zum Paradigma der Referenzialität auch ebd., 96–101 sowie Joseph Tabbi, „Locating the Literary in New Media“, in: *electronic book review* (23. September 2008), <http://www.electronicbookreview.com/thread/criticalecologies/interpretive> (letzter Aufruf 4.5.2018).

der Bedeutung mit jeweils spezifischen Handlungsperspektiven und Ressourcen, eigenen Voraussetzungen und Potenzialen“,¹³ der im Rahmen der Digitalisierung zunehmend stabilere und autoritärere Ordnungssysteme zur Produktion und Manifestation von Bedeutung ersetzt.

Entsprechend bewirken diese Verfahren gerade nicht, wie in der Montage- und Collage-Kunst der klassischen Avantgarden, ein Ausstellen der „vielen Brüche innerhalb der Moderne, ihrer Fragmentierung und Zerrissenheit“, sondern vielmehr „deren Synthese in der Gegenwart“.¹⁴ Die Verschiebung von Deutungshoheit und Autorität von einigen wenigen ‚Gatekeepern‘ hin zu unübersichtlichen und dynamischen Teilnetzwerken mit vielfältigen Akteuren schlägt sich in der digitalen Literatur – so lässt sich zumindest in Bezug auf die hier betrachteten Beispiele feststellen – in einer bedeutungsstiftenden Stabilisierung nieder. Das Ausschneiden von Elementen aus dem Meer des flüchtigen Schreibens im Internet und deren Remodellierung in konkreten Zusammenhängen ermöglicht dem Leser deren gemeinsame und vergleichende Rezeption. Diese stabilisierende Funktion zeigt sich in Bezug auf *Ox0a* explizit anhand des *Dictionary of non-notable Artists*,¹⁵ welches im Stil eines klassischen (analogen) Autorenlexikons die auf der Seite von Wikipedia zum Löschen abgelegten Künstlerbiographie-Artikel der letzten zehn Jahre inklusive der jeweiligen Begründung für deren Irrelevanz alphabetisch und nach angegebenem ‚Berufsstand‘ (*Architect, Author, Cartoonist* etc.) sortiert auflistet. Die Flüchtigkeit von Wissen im Netz sowie dessen Abhängigkeit von der kollektiven Wertung der nicht greifbaren ‚Community‘ kommen hier aufs Deutlichste zum Ausdruck und werden gleichermaßen durch die Fixierung im *Dictionary* konterkariert. Das *Dictionary* erfüllt also eine dokumentierende Funktion, indem es die flüchtigen Wikipedia-Daten in „autorisierte Knotenpunkte (nodes) im Netz [digitaler; J. N.] sprachlicher Kommunikation“¹⁶ überführt.

2.2 Paratextuelle Rahmung und Verkörperungen digitaler Textualität

Strukturen der paratextuellen Rahmung der Korpora sowie der ‚Verkörperung‘ der Texte im Zuge spezifisch elektronischer Publikationsformate ergänzen in entscheidender Weise die beschriebenen konzeptuellen intertextuellen Synthesen. Sie bilden die materiellen Charakteristika der Texte, indem sie das Resultat der

¹³Stalder (Anm. 1), 164; vgl. zur Notwendigkeit des eigenen Auswählens und Ordnen in dieser Situation auch ebd., 116f.

¹⁴Ebd., 98f.

¹⁵Gregor Weichbrodt, *Dictionary of non-notable Artists* (2016), <http://0x0a.li/de/text/dictionary-of-non-notable-artists/> (letzter Aufruf 4.5.2018).

¹⁶Christine Ivanovic, „Die Vernetzung des Textes: Im Möglichkeitsraum digitaler Literaturanalyse“, in: *Zeitschrift für digitale Geisteswissenschaften* (2017), Kap. 4, o.S., https://doi.org/10.17175/2016_010.

kulturellen und technologischen Prozesse und Bedingungen sind, welche sie als rezipierbare Publikationen instanzieren, wobei die jeweilige Ausgestaltung dieser Strukturen signifizierende Funktion im Sinne bibliographischer ‚Codierungen‘¹⁷ erlangt.¹⁸

Aufgrund der bereits dargestellten verschiedenartigen Publikationsformen bestehen für *OxOa* hierbei erweiterte Möglichkeiten der Semantisierung, da die gesamte Publikationsplattform von den Akteuren des Kollektivs selbst gestaltet ist. Raumaufteilung, farbliche Gestaltung, Schriftarten, die eigenwillige Kombination von Groß- und Kleinschreibung sowie das omnipräsente Logo bilden hier Elemente der individuellen Strukturierung als Strategie einer konsequenten literarischen Inszenierung, in welche eine variable Anzahl an Einzeltexten eingebettet ist. Die einheitsstiftende Funktion dieser Inszenierung veranschaulicht insbesondere die *These*,¹⁹ welche sich unter einem eigenen Reiter abrufen lässt. Die dort explizierte Programmatik bezieht sich gerade nicht auf einen spezifischen der präsentierten Texte. Dies wird einerseits inhaltlich deutlich, zeigt sich aber ebenso signifikant anhand ihrer Positionierung im Netzwerk der Webseite. Der eigene Reiter neben jenem, der zu den einzelnen Texten führt, sowie der Umstand, dass eine gekürzte Version der *These* ebenfalls auf jeder der Unterseiten eingeblendet wird, legen es nahe, das Grundkonzept des Kollektivs, die Erforschung der „Möglichkeiten von Text im Digitalen“, sowie dessen Definition und Verständnis von digitaler Literatur auf alle literarischen Inhalte der Seite zu beziehen. In einem an die klassischen Avantgarden gemahnenden Gestus deklariert die *These* sämtliche auf der Webseite befindliche Texte als Teile des unter dem Banner *OxOa* firmierenden „Workshop[s], Labor[s], Schaufenster[s]“ – also einer durch ihre spezifische Funktion definierten gemeinsamen Formation – und nimmt dabei gleichzeitig Einfluss auf deren theoretische und literaturgeschichtliche Einordnung.

Auch für *Tausend Tode* lassen sich aber – im Rahmen des vorgegebenen und (noch) stark von analogen Organisationsstrukturen geprägten sowie vom jeweiligen Lesegerät dependenten E-Book-Formats – spezifische paratextuelle und materielle Eigenschaften mit entscheidendem Einfluss auf die Gestalt der Publikation ausmachen. Diese beziehen sich gemäß der medial beschränkten Möglichkeiten verstärkt auf die Erscheinungsweise des textuellen Korpus selbst, unterstützen hier aber wiederum die Wahrnehmung von *Tausend Tode*

¹⁷Dieser Terminus bezieht sich auf die von McGann (in Bezug auf analoge Formate) eingeführte Differenzierung signifizierender literarischer Strukturen in „linguistic“ und „bibliographical codes“, vgl. Jerome J. McGann, *The Textual Condition*, Princeton 1991, 77.

¹⁸Materialität verstehe ich also nicht (ausschließlich) im Sinne von physikalischer Gegenständlichkeit, sondern im Sinne von Hayles an McKenzie anknüpfende Vorstellung vom „embodied text“ (Nancy K. Hayles, *My Mother was a Computer. Digital Subjects and Literary Texts*, Chicago 2005, 103; vgl. dazu auch Johanna Drucker, *What is? (a letter, writing, a word's body, a document, graphic textuality ...)* *Nine epistemological essays*, Victoria, TX, 2013, 123ff.

¹⁹Hannes Bajohr/Gregor Weichbrodt, *These*, <http://OxOa.li/de/these/> (letzter Aufruf 4.5.2018).

als kompositorische Einheit und ‚offenes Kunstwerk‘ gleichermaßen: Die versionierte Publikation bringt es mit sich, dass im Gegensatz zu analogen Auflagen oder Reihenpublikationen jede neue Version die vorausgegangene ‚überschreibt‘. *Tausend Tode* besteht also immer nur in der jeweils aktuellen Form und bildet deshalb trotz permanenter Dynamik der konstituierenden Elemente stets eine publizierte Einheit. Den Eindruck einer seriellen Publikation stützt zwar die Angabe der ‚Versionengrenzen‘ im Inhaltsverzeichnis, allerdings setzt *Tausend Tode* den vom Ausgabeprogramm vorgesehenen (analogen) Mustern des Blätterns und der Seitenzahlen, die eine lineare Lektüre befördern, eine eigene digitale Strukturierung durch die Verlinkung des Inhaltsverzeichnisses mit den einzelnen Beiträgen entgegen, welche eine veränderte Rezeptionsweise nahelegt.²⁰ Jeder der Einzeltexte kann auf diese Weise zum unmittelbaren Ausgangspunkt der Lektüre werden.

Als zentrale paratextuelle Elemente der Zusammenhangsstiftung erweisen sich die Titel der beiden Projekte: In beiden Fällen markieren sie die literarischen Erzeugnisse mehrerer Autoren als Gesamtkompositionen. In *Tausend Tode* haben die Einzeltexte überhaupt keine eigenen Titel, sondern lediglich (Kapitel-) Nummern. Der Gesamttitel stellt damit den Bezugspunkt aller Texte dar. Für *0x0a* lassen sich hingegen zwei Stufen unterscheiden: Die Einzeltexte haben jeweils eigene Titel, die auf der obersten Publikationsebene, welche hier die Webseite darstellt, unter dem Namen *0x0a* zusammengefasst sind. Dessen Wirkung als einheitsstiftende Benennung wird durch die weitere Gestaltung der Webseite als Paratext gestützt. Wie die *These* bleibt auch der Titel für jede der Unterseiten auch auf verschiedenen Lesegeräten erhalten. Er bildet also eine Konstante im Rahmen der Dynamik der Webseite und fungiert dabei gleichermaßen als Titel jenes dynamischen „Schaufensters“ und als Name des (variablen) Kollektivs der Künstler, die hier am Werk sind.²¹

Sowohl für *Tausend Tode* als auch für *0x0a* gewinnen die Haupttitel an Relevanz, da die unter diesen ‚Labels‘ publizierenden Künstler wechseln. Die Titel repräsentieren hingegen in einer komplexen Verschaltung von Signifikationsebenen die gemeinsamen Referenzrahmen der Kompositionen, indem sie an die Stelle des einheitsstiftenden Autornamens treten, dabei jeweils das vereinende Konzept der Einzeltexte explizieren und zusätzlich selbst-reflexiv auf deren spezifische Produktions- und Distributionspraktiken mithilfe

²⁰Vgl. hierzu Hayles (Anm. 18), 90f.: „[N]avigational functionalities are not merely ways to access the work but part of a work’s signifying structure. An encyclopedia signifies differently than does a realistic novel in part because its navigational functionalities anticipate and structure different reading patterns“.

²¹Das Zentrum des Kollektivs bilden die beiden prominent in der *These* genannten Autoren Gregor Weichbrodt und Hannes Bajohr. Die Texte entstehen allerdings (wie in den jeweils vorgeschalteten Einleitungen ausgewiesen) z.T. in Kollaboration der beiden Künstler und mit anderen Künstlern. Unter den Texten befindet sich zudem (bisher) ein Text: *LENZ* von Dagmar Kraus (2016), <http://0x0a.li/de/text/lenz/> (letzter Aufruf 4.5.2018). Deshalb gehe ich von einer grundsätzlichen Variabilität der Beteiligten aus.

digitaler Programme und Kommunikationskanäle verweisen: So benennt der Titel *Tausend Tode schreiben* nicht nur das gemeinsame Thema der Texte, sondern das Verb ‚schreiben‘ im Präsens bringt zudem den Status des Bandes als ‚Work in Progress‘ zum Ausdruck, wobei einerseits die Tätigkeit der verschiedenen Autoren angesprochen ist, andererseits aber auch die Interpretation einer ‚Selbstbeschreibung‘ der Tode nahegelegt wird, was wiederum die Autoren in den Hintergrund treten lässt.

0x0a demonstriert als Hexadezimalcode für den Zeilenumbruch die Differenz der Texte zu den Praktiken der analogen Literaturproduktion und antizipiert das Moment der Maschinenlesbarkeit, welches das den Texten gemeinsame Verfahren ermöglicht.²² Indem das Logo auf der Webseite mit einem dynamisierenden Algorithmus unterlegt ist,²³ wird der nur visuell repräsentierte „broken code“²⁴ auf anderer Ebene wieder zum ‚laufenden‘, weil ausgeführten Code. Das Logo verkörpert so die Variabilität auf den beiden durch *0x0a* benannten Ebenen der Texte und der Künstler und verweist dabei gleichzeitig auf die verschiedenen Ebenen digitaler Schrift zwischen Binärcode und Oberflächentext,²⁵ die deren spezifische und von der stabilen Druckschrift konstitutiv verschiedene Funktionsweise ausmachen.²⁶ Der beim Überfahren mit dem Cursor angezeigte ‚Untertitel‘ „Eskaliertes Schreibenlassen“ thematisiert zudem analog und gleichzeitig komplementär zum Titel *Tausend Tode schreiben* die (Nicht-)Schreibpraxis des Kollektivs.

Die Einbettung der Einzeltexte in Korpus-übergreifende materielle und paratextuelle Strukturen trägt maßgeblich zur Konstituierung kompositioneller Einheiten bei. Dabei stabilisieren sie das bedeutungsstiftende Referenznetzwerk durch eine dynamische Begrenzung, die jeweils in Verbindung zu einem spezifischen Konzept steht.²⁷ Die Einzeltexte von *Tausend Tode* und *0x0a* erscheinen durch diese Verkörperungen aus der „Masse der in einer Gemeinschaft umlaufenden sprachlichen [bzw. hier eher ‚schriftlichen‘; J. N.] Mitteilungen“ in sozialen Netzwerken,

²² Entgegen den programmatischen Aussagen auf der Homepage des Kollektivs lässt sich ‚0x0a‘ sehr wohl aussprechen und analog darstellen, wobei die Buchstaben-Zahlen-Kombination aber – wie das Logo selbst demonstriert – eben nicht denselben Zweck erfüllt wie als maschinenlesbarer Code.

²³ Der Schriftzug ‚0x0a‘ wird im Abstand von zwanzig Sekunden von verschiedenen Buchstaben ‚durchlaufen‘, deren genaue Kombination mit bloßem Auge nicht auszumachen ist, bevor er wieder bei der Ausgangskombination anhält. Der Blick in den Quellcode verrät, dass hierfür ein Algorithmus jeweils zufällig zehn Buchstaben aus dem Alphabet pro Buchstabe und Ziffer auswählt.

²⁴ John Cayley, „The Code is not the Text (unless it is the Text)“, in: *electronic book review* (10. September 2002), <http://www.electronicbookreview.com/thread/electropoetics/literal> (letzter Aufruf 4.5.2018).

²⁵ Vgl. ebd.

²⁶ Vgl. Hayles (Anm. 18), 100f.

²⁷ Diese Stabilisierung repräsentieren wiederum die gegenüber ihrem dynamischen Umfeld stabilen Titelstrukturen.

Kommentarbereichen etc. herausgehoben.²⁸ Diese Stabilisierung verleiht ihnen – so könnte man mit Lotman argumentieren – Textstatus,²⁹ ohne dabei jedoch ihr dynamisches Verhältnis zu anderen medialen Äußerungen gänzlich aufzuheben.

3 Vervielfältigte Autorschaft und Dispersion der Autorfunktionen

Diese einheitsstiftende materielle Verkörperung steht, wie bereits angedeutet, im Kontrast zur Vervielfältigung der Autorschaft in beiden Korpora. Digitale Autorschaft ist durch Kollaboration, Kompilation und Unabgeschlossenheit geprägt und repräsentiert auf diese Weise zentrale Elemente der Kultur der Digitalität. *Tausend Tode* und *OxOa* repräsentieren zwei verschiedene Spielarten dieser Zusammenarbeit, die an die unterschiedlichen Produktionspraktiken gebunden sind, aber gleichermaßen zu Diversifizierungen und Verschränkungen der Kategorien ‚Urheber‘ und ‚Schöpfer‘³⁰ führen: *Tausend Tode* kennzeichnet eine akkumulative Kooperation, d.h. die einzelnen Autoren produzieren ihre Texte grundsätzlich unabhängig voneinander. Aufgrund der versionierten Publikation entstehen aber alle Texte, die im Anschluss an Version 1 hinzukommen, im Bewusstsein der bereits bestehenden ‚Kapitel‘ und fließen im Moment der Publikation in denselben paratextuellen Rahmen ein. So werden sie Teil der Verkörperung des vom Verlag gesetzten Gesamtkonzepts, das gerade nicht individuell ist, sondern vielmehr den gemeinsamen Referenzpunkt der akkumulierten Texte darstellt. Hierbei handelt es sich also um eine projektspezifische Kooperation, an der – begünstigt durch die medialen Bedingungen – eine Vielzahl von Autoren beteiligt ist, ohne dass hierfür Rücksprachen zwischen den einzelnen Verfassern notwendig sind.³¹

²⁸ Juri M. Lotman: *Kunst als Sprache. Untersuchungen zum Zeichencharakter von Literatur und Kunst*, Leipzig 1981, 35.

²⁹ Lotman (ebd.) definiert einen ‚kulturbezogenen Textbegriff‘ in Abgrenzung zu einem rein linguistischen: „Demzufolge wird die gesamte Masse der in einer Gemeinschaft umlaufenden sprachlichen Mitteilungen aufgefaßt als Nichttexte, vor deren Hintergrund sich eine Gruppe von Texten abhebt, die Merkmale einer zusätzlichen, im System der betreffenden Kultur signifikanten Realisierung aufweisen.“

³⁰ Vgl. zu dieser Unterscheidung u.a. Matías Martínez, „Autorschaft und Intertextualität“, in: Ders./Fotis Jannidis/Gerhard Lauer u.a. (Hg.), *Die Rückkehr des Autors. Zur Erneuerung eines umstrittenen Begriffs*, Tübingen 1999, 474.

³¹ Im Gegensatz dazu basieren kollaborative Schreibprojekte wie Fan Fictions und Mitschreibeprojekte im Netz auf einer stärkeren Interaktion zwischen den einzelnen Autoren bzw. einem narrativen Zusammenhang der einzelnen Texte (vgl. dazu den Beitrag von Thomas Ernst in diesem Band). Hierbei spielt allerdings ebenfalls teilweise die Initiation und/oder Präsentation durch eine moderierende Instanz eine Rolle (vgl. z.B. das vom Penguin-Verlag begonnene und verantwortete Projekt *A Million Penguins*). Dies gilt auch für *Beim Bäcker* (1996–2000), initiiert von Claudia Klinger, das zudem in seiner Organisation nach mit den Namen der Verfasser benannten Kapiteln, die einen losen thematischen Zusammenhang aufweisen, wie *Tausend Tode* [Frohmann [Anm. 3)] eher kooperativ funktioniert. Neben einer (abgestuften) Differenzierung

In den Texten von *Ox0a* findet maßgeblich ‚Sprachmaterial‘ anderer Autoren Verwendung, wodurch bereits die Einzeltexte in den meisten Fällen polyphon sind. Zentrales Moment der Textproduktion ist bei *Ox0a* nicht das Schreiben der publizierten Texte, sondern vielmehr das Schreiben der technischen Skripte, die anhand der ausgewählten Korpora und der gesetzten Bedingungen Such-, Selektions-, Sortier- und Kombinationsaktionen ausführen und so die einzelnen Texte gemäß dem gesetzten Konzept erzeugen. Diese neue Ebene des Schreibens von digitaler Literatur thematisieren die den Einzeltexten von *Ox0a* jeweils vorgeschalteten Erklärungen, in denen dargestellt wird, welche Programme hier wofür geschrieben und eingesetzt wurden. Diese mit auktorialen Vorworten vergleichbaren Texte zeugen von der künstlerischen Relevanz des technischen Schreibens für die Literatur von *Ox0a*, die sich ebenfalls in der Gestaltung des paratextuellen Rahmens der Webseite niederschlägt. Die Erklärungen des Kollektivs inszenieren zum einen Autorschaft als kreativen und technischen Produktionsprozess und machen letzteren dabei zum anderen transparent. Dieser Transparenz entspricht auch die begleitende Publikation des für die ‚Samples‘ verwendeten Gesamtkorpus, etwa bei *Glaube, Liebe, Hoffnung*.³²

Autorschaft wird in ihren Ausprägungen von Urheberschaft und Schöpfer-tum deshalb zwar disparat, aber keinesfalls obsolet.³³ Für *Ox0a* sind prinzipiell verschiedene Kompetenzen zwischen programmiertechnischem Know-how, der Kenntnis der vielfältigen Diskurse, auf die sich die Texte beziehen, und kreativen Überlegungen zur Konzeptionierung der Samples gefragt. Auf deren Basis entsteht

in kooperative und kollaborative Projekte gilt es daher, produktionsästhetisch verschiedene Abstufungen der (De-)Hierarchisierung und der Sichtbarkeit von Autorschaft im Rahmen gemeinschaftlicher digitaler Literaturproduktion zu unterscheiden (vgl. die Differenzierung bei Florian Hartling, *Der digitale Autor. Autorschaft im Zeitalter des Internets*, Bielefeld 2009, 267; zur Autorschaft und zum Autor-Leser-Verhältnis in dem vom Penguin-Verlag initiierten Projekt *A Million Penguins* (2007) ausführlich Zimmermann (Anm. 5), 198–213 sowie insgesamt zu diesem Thema auch Julia Nantke, „Multiple Autorschaft als digitales Paradigma und dessen Auswirkungen auf den Werkbegriff“, in: Svetlana Efimova (Hg.), *Textpraxis. Digitales Journal für Philologie* 2 (2018), Sonderausgabe #3, <https://doi.org/10.17879/77159516645>.

³²Nicht nur die algorithmische Produktionsweise, sondern auch diese Form der Offenlegung von Werkzeugen und Verfahren, mit denen bestehende Korpora be- und verarbeitet werden, sowie das Zurverfügungstellen der Gesamtkorpora weisen Parallelen zu den Arbeitstechniken der digitalen Literaturwissenschaften auf. Der Einsatz dieser Techniken erfolgt hier allerdings unter einer gänzlich anderen, nämlich künstlerischen Stoßrichtung, vgl. Annette Gilbert, „Möglichkeiten von Text im Digitalen“. Ästhetische Urbarmachung von korpuslinguistischen Analysetools in der generativen Literatur der Gegenwart am Beispiel des Textkollektivs *Ox0a*“, in: *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* 91/203 (2017), <https://doi.org/10.1007/s41245-017-0038-y>. Diese zielt nicht auf Erkenntnisgewinn und Reliabilität, sondern auf die kreative Be- und Ausnutzung des Korpus. (Diese Überlegungen erfolgten im Anschluss an Gespräche im Rahmen des Symposiums u.a. mit Peer Trilcke, dem ich für Anregungen in diese Richtung sowie den Literaturhinweis danke.)

³³Vgl. Hartling (Anm. 31), 10.

in der Interaktion von Mensch und Maschine, von ‚fremden Stimmen‘ und rekonzeptualisierender Modellierung zunächst jeder einzelne Text, der mit seinen inkorporierten Textfragmenten letztlich in den „konzeptuellen Rahmen“³⁴ unter dem gemeinsamen Label *Ox0a* eingeht. Rezipientenseitig lässt sich erst hier „das Konzept in seiner umfassenden Gestalt ablesen, deren Umfang dem Erfinder oder der Erfinderin selbst nicht bewusst gewesen sein muss“,³⁵ wodurch das Korpus zum Objekt von Interpretationsprozessen werden kann.

Auch digitale Literatur hat also Autoren; die Ausgestaltung digitaler Texte vollzieht sich allerdings verstärkt im Rahmen variabler Gruppierungen von (beabsichtigt oder unbeabsichtigt) am Prozess der konzeptionellen, konkret-textuellen, technischen und materiellen Erzeugung beteiligter Akteure. *Tausend Tode* und *Ox0a* zeigen, dass hierbei für das Feld digitaler Literatur nicht eindeutig zwischen spezifischen Zuständigkeiten im Sinne von ‚Schöpfern‘ und ‚Urhebern‘ zu unterscheiden ist, sondern vielmehr verschiedene Möglichkeiten der Verteilung und des Zusammenfallens auktorialer Funktionen bestehen.³⁶

Die von Foucault herausgestellte Funktion der Gruppierung und „In-Beziehung-Setzen der Texte untereinander“³⁷ erfüllen für die beiden hier diskutierten Beispiele nicht die Autoren bzw. deren Namen, sondern maßgeblich die Titel in Wechselwirkung mit der Einbettung der Texte in ihre Publikationsformate.³⁸ Ein von einem Verlag gesetztes Thema kann hierbei ebenso zum konzeptuellen Ausgangs- und Bezugspunkt einer Publikation werden wie das *Hohelied der Hoffnung* (bei *Glaube, Liebe, Hoffnung*) oder der literarische Text eines anderen Autors (z.B. bei *LENZ*³⁹). Auch das Sampling verweist als gemeinsames Verfahren der Textproduktion von *Ox0a* auf eine Autorfunktion, wobei technische Fähigkeiten die Möglichkeiten der Literaturproduktion determinieren.

³⁴ Carlos Spoerhase, „Was ist ein Werk? Über philologische Werkfunktionen“, in: *Scientia Poetica* 11 (2007), 290.

³⁵ Bajohr (Anm. 2), 16.

³⁶ Vgl. den Hinweis hierauf bereits in Simone Winko, „Lost in Hypertext? Autorkonzepte und neue Medien“, in: Dies./Fotis Jannidis/Gerhard Lauer u.a.: *Die Rückkehr des Autors. Zur Erneuerung eines umstrittenen Begriffs*, Tübingen 1999, 530, sowie in handlungstheoretischer Perspektive Hartling (Anm. 31).

³⁷ Michel Foucault, „Was ist ein Autor?“, in: Ders., *Schriften zur Literatur*, hg. von Daniel Defert/François Ewald, Frankfurt a. M. 2003, 244.

³⁸ Dies gilt noch verstärkt für Mitschreibe-Projekte im Netz, die einen narrativen Zusammenhang bilden, wie bspw. die beiden ‚interaktiven Netzmomane‘ *Magische Welt. Íja Macár* (1999–2008), initiiert von Roger Nelke, http://www.drachental.de/ijamacar/im_index.htm, und *Die Perlen von Caala-Elen* (1999–2008), initiiert von Susann Ulshöfer, <https://web.archive.org/web/20070314051902/http://www.zauberfee.de:80/zauberbuch/caala-elen/index.htm> (letzter Aufruf der Links 4.5.2018). Die Titel und die jeweilige Webseite bilden die zentralen Referenzpunkte der Texte; individuelle Autorschaften sind allerdings ebenfalls zu Beginn der einzelnen Kapitel und in der Übersicht ausgewiesen.

³⁹ Kraus (Anm. 21).

Diese einheitsstiftenden Momente stehen in beiden Beispielen allerdings in permanenter Spannung zur Dispersion auktorialer Signale in Stil, Sprache und Inhalt, was weitere notwendige Differenzierungen bedingt: Autorschaft bezieht sich auf ganze Texte oder auf die kombinierten Einzeltexte oder Textfragmente, geschrieben von professionellen und Laienautoren, deren Schreiben sich im Lichte des gesetzten Konzepts oder mit völlig anderer Zielrichtung vollzieht und letzterem erst nachträglich einverleibt wird. Dennoch erscheinen auch für *Ox0a* nicht nur die Tätigkeiten der aktuellen Autoren für die Wirkung der digitalen Texte relevant, sondern häufig auch die der ‚unfreiwilligen‘ Autoren, deren jeweils spezifischer Status bzw. Zugehörigkeit zu einem spezifischen Referenzfeld (Autoren des Kanons für *Durchschnitt*,⁴⁰ Facebook-Community von Pegida für *Glaube, Liebe, Hoffnung*) sich entscheidend auf die Semantisierung der Texte auswirkt.

Autorschaft funktioniert nicht mehr als Garant für die „Einheit des Schreibens“.⁴¹ Stattdessen sind verschiedene ‚textuelle Handlungen‘ hinsichtlich der unterschiedlichen in ihnen zum Ausdruck kommenden Autorfunktionen jeweils individuell zu differenzieren und in ihren Interaktionen zu beschreiben, da sich auf diese Weise Aussagen über das Zusammenwirken von Schreib-, Kompositions-, Kompilations- und Referenzstiftungsstrategien im jeweiligen Korpus treffen lassen.

4 Werkhaftigkeit als digitale Kommunikationsstrategie

Die Vorstellung des in sich (ab)geschlossenen und stabilen Textes⁴² mit eindeutig zuordenbarer Autorschaft, der sich im Buch als „immutable mobile“⁴³ realisiert, verliert für die untersuchten Beispiele an Relevanz. Als Garanten für die textuelle

⁴⁰ Hannes Bajohr, *Durchschnitt* (2015), <http://Ox0a.li/de/text/durchschnitt/> (letzter Aufruf 4.5.2018).

⁴¹ Foucault (Anm. 37), 249. Dies gilt auf sprachlicher und stilistischer Ebene. Zu den komplexen Implikationen vielfältiger Autorschaft für urheberrechtliche Fragen vgl. den Beitrag von Thomas Ernst in diesem Band und insbesondere zu den damit verbundenen Verantwortlichkeiten auch Gilbert (Anm. 32).

⁴² U.a. Roland Reuß, „Text, Entwurf, Werk“, in: *Text. Kritische Beiträge* 10 (2005), 7, und Spoerhase (Anm. 34), 289, sehen gerade in der Stabilität des Geschriebenen das entscheidende Kriterium für dessen Textstatus (im Gegensatz zum Entwurf), der nach Spoerhases tendenziell produktionsästhetischem Verständnis wiederum die Voraussetzung für dessen „Werkwerdung“ bildet (vgl. ebd.). Reuß’ eher rezeptionsästhetischem Werkverständnis folgend, können hingegen in Abhängigkeit von literaturgeschichtlichen Zuschreibungen auch „Entwürfe [...] in einen Werkzusammenhang transzendieren“ (ebd., 10).

⁴³ Bruno Latour, „Drawing Things Together“, in: Michael Lynch/Steve Woolgar (Hg.), *Representation in Scientific Practice*, Cambridge, MA, 1990, 26. Zur Abhängigkeit dieser Stabilität von sozialen und kulturellen Konstruktionen vgl. Lisa Gitelman, *Paper Knowledge. Toward a Media History of Documents*, Durham/London 2014, 113.

(Werk-)Einheit können aufgrund der verschiedenen Formen kompilativer Textproduktion zudem weder *der* Autor noch die konventionell hierfür veranschlagten Textebenen der Handlung, des Stils oder der Sprache⁴⁴ herangezogen werden.

In systematischer Spannung zur konstitutiven Unabgeschlossenheit und auktorialen Vielfalt der Korpora stehen allerdings die dezidierte konzeptionelle Einfassung und die Einbettung in (spezifisch elektronische) Publikationsformate als zentrale Merkmale von Werkhaftigkeit.⁴⁵ Verbundenheit konstituiert sich in beiden Beispielen nicht durch auktorial verfügte oder am Text ablesbare Geschlossenheit, sondern vielmehr anhand gemeinsamer konzeptioneller, struktureller, paratextueller und/oder inhaltlicher Parameter, die eine dynamische Formation variabler, aber stets aufeinander bezogener Positionen, Äußerungen und Akteure definieren.⁴⁶ *Tausend Tode* und *Ox0a* zeugen damit von einer grundsätzlichen Verschiebung im Verhältnis von Autorschaft, Text und Werk: Nicht mehr der einzelne Autor personifiziert die materielle Arbeit an einem Text bzw. auf diesen hin mit seiner Schaffenszeit. Deshalb existiert ebenfalls keine singuläre Instanz mehr, die als individueller geistiger „Schöpfer“ des (immateriellen) Werks gelten kann.⁴⁷

Indem für *Tausend Tode* und *Ox0a* das jeweilige gemeinsame Konzept, die konkreten textuellen Realisierungen sowie die materielle Verkörperung in wechselseitiger Abhängigkeit die Wirkung der Korpora als „übergreifende ästhetische Kommunikationseinheit“⁴⁸ hervorbringen, fallen Kriterien für materielle Textualität und ideelle Werkhaftigkeit zusammen: Das Werk entsteht als ein sich unmittelbar materiell realisierendes und dabei interaktiv konzeptionell ausgestaltetes im Prozess des Zusammen-Schreibens (von Text und/oder durch mehrere Autoren).⁴⁹ Die Einzeltexte stellen dabei gleichzeitig individuelle

⁴⁴Neben den durch die verschiedenen ‚Stimmen‘ verursachten permanenten Duktus-Wechseln zeichnen sich beide Beispiele durch eine Mischung deutscher und englischer Texte aus. Bei *Ox0a* besteht zudem für die Paratexte der Webseite und z.B. auch für die *These* die Möglichkeit, zwischen Deutsch und Englisch zu wählen.

⁴⁵Spoerhase nennt als einschlägige Kriterien für den Werkstatus eines Textes den Titel, den Veröffentlichungsakt, die Autorabsicht sowie den Geschlossenheits- bzw. Vollendungsgrad: Ders. (Anm. 34), 288.

⁴⁶Letztlich erweist sich auch die vermeintliche Stabilität analoger Texte als kulturelle Konstruktion, die maßgeblich durch das Medium Buch und die Vorstellung einer (auch retrospektiv erfassbaren) Autorintentionalität forciert wird. Davon zeugen viele *Digitale Editionen*, welche die Stabilisierung nachträglich durch die parallele Präsentation verschiedener Fassungen redynamisieren. Grundlegende Unterschiede zu den hier diskutierten digitalen Texten bestehen aber dennoch in Bezug auf das Verhältnis von Produktion und Rezeption und die transportierten Vorstellungen von Autorschaft.

⁴⁷Vgl. exemplarisch zur Trennung von Text als materiellem Träger, der einen Urheber hat, und Werk als Rahmen, der einem „konzeptuellen Schöpfer“ zugerechnet wird: Martínez (Anm. 30), 474f.

⁴⁸Spoerhase (Anm. 34), 307.

⁴⁹Ähnlich beschreibt Simanowski das „Zugleich der beiden Betrachtungsebenen Konzept und Inhalt“ als „Wesensmerkmal vieler kollaborativer Schreibprojekte“: Der. (Anm. 8), 45.

Manifestationen der kollektiven Idee und Elemente des übergeordneten Programms dar. Ihre fortgesetzte Publikation als Teile der jeweils ‚überschriebenen‘ Einheit bewirkt die stetige Aktualisierung der Gesamtkomposition auf materieller und semantischer Ebene. Die Maschinenlesbarkeit der Korpora und im Fall von *OxOa* deren freie Verfügbarkeit im Netz⁵⁰ bergen zudem immer das Potenzial des ‚Zurückfließens‘ der Texte in den Kommunikationsstrom bzw. für deren Verwendung in anderen Ko- und Kontexten.⁵¹

Das Ineinandergreifen von Konzept und Realisierung kann als Spezifikum der Internet-Kultur gedeutet werden, deren mediale Praxis sich grundsätzlich durch die Unmittelbarkeit von Publikationsprozessen auszeichnet. Das digitale Paradigma ermöglicht einerseits die kostengünstigere „institutionelle Rahmung“⁵² durch E-Book und Print-on-Demand, andererseits tritt die Notwendigkeit einer solchen offiziellen Sanktionierung zugunsten der technischen und kreativen Fähigkeiten der Künstler bei der Gestaltung einer eigenen Webpräsenz in den Hintergrund. Im Hinblick auf die Vielzahl und Vielfalt der digital umlaufenden Äußerungen fungiert dieses Ineinandergreifen aber, wie dargestellt, auch im Sinne einer semantisierenden Markierung der Texte durch Zusammenfassung in einer signifikanten Beobachtungseinheit.

Dies verweist auf eine geänderte Kommunikationssituation im Rahmen digitaler Publikation: Lesbarkeit ist im Internet bzw. mithilfe digitaler Leseprogramme schnell und kostengünstig erreicht, Aufmerksamkeit im Sinne tatsächlicher Lektüren garantiert diese aber noch lange nicht. Die gewandelte ‚institutionelle‘ Rahmung durch konzeptuelle Einfassung und exponierte Präsentation erweist sich also wiederum als diskursive Strategie, welche die Texte von der Masse flüchtiger digitaler Kommunikation abhebt. Die permanente Aktualisierung kann im Rahmen dieser Strategie ebenfalls als entscheidender Faktor gelten, der auf die gewandelten Mechanismen in der digitalen Umwelt reagiert: Referenzialität erlangt nach Stalder in der digitalen Kommunikation „performativen Charakter“, denn „[w]as nicht dauernd verwendet und erneuert wird, verschwindet“.⁵³ Tausend Tode stirbt in der digitalen Literatur nicht nur *der* Autor, sondern ebenso ‚sterben‘ tausende digital publizierte Texte, die aufgrund mangelnder Exponiertheit nicht (mehr) gelesen werden. Die stetig aktualisierte

⁵⁰Neben der freien Verfügbarkeit und der gesteigerten Frequenz der De- und Rekontextualisierung von Texten im Netz sind auch „die allgemeine Verfügbarkeit und Zugänglichkeit der entsprechenden Tools [zu deren Bearbeitung; J. N.] und ihr Siegeszug in allen Bereichen des täglichen Lebens“ spezifisch für die Kultur der Digitalität und Differenzmerkmale zu „der ersten Welle generativer Literatur der 1960er Jahre“, siehe Gilbert (Anm. 32).

⁵¹Die Texte können also tatsächlich mehrere Werke „durchqueren“, vgl. Roland Barthes, „Vom Werk zum Text“ (1971), in: Stephan Kammer/Roger Lüdeke (Hg.), *Texte zur Theorie des Textes*, Stuttgart 2005, 42; die Werke sind dabei aber selbst dynamische Einheiten und nicht im Sinne von Barthes’ „bruchstückhafter Substanz“ (ebd.) als stabile materielle Schwundstufe der Texte zu begreifen.

⁵²Spoerhase (Anm. 34), 287.

⁵³Stalder (Anm. 1), 128.

Akkumulation und konzeptionelle Einbindung befördern also auch Werkhaftigkeit im Sinne der von Reuß hierfür veranschlagten „Wirkungsgeschichte“.⁵⁴

5 Rezeptionsstrategien: Fragmentieren, Vernetzen, Mitschreiben

Durch ihre Unabgeschlossenheit und ihren (wachsenden) Umfang erscheinen *Tausend Tode* und *Ox0a* als „Produktion[en] ohne Produkt“, also mit Barthes als „schreibbare[] Text[e]“.⁵⁵ Nicht mehr eine auf Vollständigkeit bedachte, durchgehende Lektüre von Anfang bis Ende ist gegenüber diesen Korpora die angemessene Lektürehaltung. Sie verunmöglichen die reine „Konsumierung“⁵⁶ und forcieren eine rekursive und fragmentarische Rezeptionsweise. Ihre jeweils Einheit stiftenden referenziellen Bezugsfelder und die Kontrastierung der Linearität der Textabfolge durch die Linkstruktur der Inhaltsverzeichnisse regen zu individuellen Verknüpfungen zwischen ihren einzelnen Teilen an. Diese können im Fall von *Ox0a* zwischen den Texten des Korpus, den jeweils in Paratexten ausgeführten Konzepten ihrer Konstruktion sowie zwischen den jeweiligen digitalen ‚Überarbeitungen‘ und ihren (literarischen) Vorlagen erfolgen.

Die Relation von Gesamtkonzept und individuellen Einzeltexten sowie die unmittelbaren Wechselwirkungen zwischen beiden Ebenen legen verschiedene Rezeptionsstrategien zwischen übergreifender Synopse und fokussierendem ‚Close Up‘ nahe, welche sich stets im Spannungsfeld von strukturell vorgegebenen Mustern und individuellen Interventionen bewegen.⁵⁷

So erweisen sich die Texte von *Ox0a* insbesondere in der Zusammenschau in ihren vielfältigen Facetten zwischen trivialer Alltagskultur und gewichtigem

⁵⁴ Reuß (Anm. 42), 10. Inwieweit diese Strategien allerdings als spezifisch digitale Markierungen von Literarizität im Diskurs funktionieren, d.h. tatsächlich zu einer rezipientenseitigen Akzeptanz der neuen ‚Rahmenbedingungen‘ beitragen, ist aktuell angesichts des Fehlens verbindlicher literarischer Strukturen im digitalen Raum schwer zu beurteilen.

⁵⁵ Roland Barthes, *S/Z*, Frankfurt a. M. 1976, 9. Dies gilt allerdings in etwas anderer Form als von Barthes (für die Dynamik analoger Literatur) veranschlagt. Für Barthes ist der „plurale Text“, dessen höchste/vollständigste Stufe der schreibbare Text darstellt, „immergleich und neu“ zugleich (ebd., 21), während die digitalen Korpora tatsächlichen Veränderungen unterliegen, die sich auf die pluralen Lektüre- und Deutungsmöglichkeiten des Rezipienten auswirken.

⁵⁶ Ebd., 20; vgl. dazu auch Barthes (Anm. 51), 48.

⁵⁷ Vgl. zu diesem Verständnis von Barthes' Konzept des schreibbaren Textes in Abgrenzung zu George P. Landow, *Hypertext 3.0. Critical Theory and New Media in an Era of Globalization*, Baltimore 2006, auch Terry Harpold, *Ex-foliations. Reading Machines and the Upgrade Path*, Minneapolis 2009, 150f. und 172. Roberto Simanowski spricht in Bezug auf die von ihm diskutierten Beispiele digitaler Literatur in ähnlichem Sinne von einer „Verdopplung der Rezeptionshaltung“, siehe: Ders., „Autorschaft und digitale Medien. Eine unvollständige Phänomenologie“, in: Lucas Marco Gisi, Urs Meyer u.a. (Hg.), *Medien der Autorschaft. Formen literarischer (Selbst-)Inszenierung von Brief und Tagebuch bis Interview und Fotografie*, München 2013, 254.

Wertediskurs als Bestandsaufnahme aktueller kultureller und gesellschaftlicher Tendenzen und Befindlichkeiten, wie sie sich im Internet als quasi unbegrenzte „Weltbildmaschine“⁵⁸ präsentieren. Diese äußern sich maßgeblich durch die gesetzten Themen, aber – wie insbesondere *Glaube, Liebe, Hoffnung* zeigt – signifikant auch in spezifischen sprachlichen Formationen, die nur in der genauen Lektüre des Einzeltextes zu erfassen sind.⁵⁹

Die vergleichende Gegenüberstellung einzelner Texte und Konzepte von *Ox0a* eröffnet wiederum spezifische Reflexionsräume: Die parallele Betrachtung des *Dictionary of non-notable Artists* und des ‚Romans‘ *Durchschnitt* offenbart in der Relationierung der Konzepte im Verhältnis zu den durch sie restrukturierten Korpora die unterschiedlichen Funktionen und Wirkungen des Sampling in Abhängigkeit von den jeweiligen Ausgangstexten, wobei die Machtmechanismen verschiedener digitaler und analoger Ordnungssysteme bei der Zuschreibung von kultureller Relevanz zum zentralen Thema des Vergleichs avancieren. Während das Verfahren in Bezug auf die digitale (Un-)Ordnung eine stabilisierende Funktion erfüllt, lässt sich für dessen Anwendung auf ein nach den Parametern des traditionellen Literaturbetriebs organisiertes Korpus gerade das Gegenteil feststellen: Für *Durchschnitt* „wurden alle Bücher aus *Der Kanon. Die deutsche Literatur: Romane*, herausgegeben von Marcel Reich-Ranicki [...], als Textkorpus verwendet, mit Python dessen durchschnittliche Satzlänge bestimmt (18 Wörter), alle Sätze anderer Länge aussortiert und das Ergebnis anschließend alphabetisch geordnet“.⁶⁰ Das Ergebnis ist ein Durchschnitt durch den Kanon, der das emphatische, auf Einzigartigkeit gerichtete Werkverständnis Reich-Ranickis ad absurdum führt, indem das algorithmische Verfahren anhand des ‚profanen‘ Parameters der Satzlänge gerade auf die Durchschnittlichkeit der Romane zielt.⁶¹

⁵⁸ Stefan Schmitt, „Mehr Aber als Ja“, in: *DIE ZEIT* 44 (29. Oktober 2015).

⁵⁹ U.a. das Projekt *dadaOverload* (2016) des Künstlerkollektivs *AND-OR* (<http://www.and-or.ch/dadaoverload/>, letzter Aufruf 4.5.2018), welches ebenfalls algorithmusbasiert funktioniert und mit der Verwendung von Twitter-Posts den ‚Output‘ digitaler Kommunikation mit literarischen Avantgarde-Texten konfrontiert, verweist im Vergleich mit den Texten von *Ox0a* darauf, dass hier wiederum zwischen verschiedenen Graden der Lesbarkeit differenziert werden muss, die nicht mit Unlesbarkeit gleichzusetzen sind, sondern vielmehr spezifische (Grenz-)Formen der Leseerfahrung bewirken (vgl. in diesem Zusammenhang auch Löss Pequeño Glazier, *White faced bromeliads on 20 hectares* (1999), http://collection.eliterature.org/1/works/glazier__white-faced_bromeliads_on_20_hectares.html (letzter Aufruf 4.5.2018). Im Gegenzug sind die Dynamik des Textes und der Aktivitätsgrad des Rezipienten in Bezug auf *dadaOverload* gesteigert, indem dieser am ‚Spiel‘ der De- und Rekomposition teilnehmen und selbst Textschnipsel zu dem sich permanent verändernden textuellen Gebilde beisteuern kann.

⁶⁰ <http://Ox0a.li/de/text/durchschnitt/> (letzter Aufruf 4.5.2018).

⁶¹ Auch dieses Verfahren stellt eine kreative ‚Umnutzung‘ quantitativer Analysemethoden der *Digitalen Literaturwissenschaft* dar und der Titel *Durchschnitt* kann in diesem Sinne auch als Hinweis auf die statistische Fokussierung auf Durchschnittswerte zur Bestimmung grundlegender Texteigenschaften großer Korpora verstanden werden. Gilbert verweist darauf, dass sich unter dieser Perspektive *Durchschnitt* „auch als Versuch lesen [lässt], auf das von Reich-Ranicki selbst benannte Problem der unzureichenden Kenntnis und Lektüre des Kanons eine überraschende, gleichermaßen verstörend aporetische wie überzeugend zeitgemäße Antwort zu finden

Dabei treten durch die alphabetische Ordnung auch sprachliche ‚Durchschnittlichkeiten‘ wie eine Vielzahl gleichlautender Satzanfänge zu Tage, die Ansatzpunkte für einen Vergleich der hier gesampelten literarischen ‚Stimmen‘ bieten.⁶²

Auch im Hinblick auf *Tausend Tode* erzeugt gerade die Heterogenität der ‚Stimmen‘, Genres und Zugriffe auf das Thema ‚Tod‘ den anvisierten überpersonellen ‚Metatext‘. Die einzelne Stimme verliert im Rahmen des Ganzen an Gewicht, steht aber im Moment des *Close Reading*, also wenn der Rezipient ihr seine Aufmerksamkeit zuteil werden lässt, wiederum für sich als subjektive Artikulation der mit der menschlichen Grundkonstante ‚Tod‘ verbundenen Erfahrungen, Vorstellungen und Deutungen. Je dichter sich der Rezipient einem der Einzeltexte annähert, desto spezifischer, konkreter, individueller artikuliert sich die Darstellung des Todes. *Tausend Tode* ist damit ein ‚Abbild‘ des menschlichen Verhältnisses zum Tod, der in vielen Versionen allgegenwärtig ist und sich doch nur aus der Nähe individualisiert und konkretisiert.

Indem der thematische Referenzrahmen von *Tausend Tode* deutlich enger ist als jener von *Ox0a*, ermöglicht das Werk – vermittelt durch seine digitalen Produktions- und Rezeptionsbedingungen – über die Textauswahl hinaus spezifische Formen des leserseitigen Agierens in Bezug auf den Gesamttext: Die vom E-Book-Reader vorgesehene Suchfunktion kann zur Erzeugung individueller intertextueller Verknüpfungen eingesetzt werden, die wiederum die selektive Lektüre befördern. Im Prinzip ist eine solche individuelle Intertextualität für alle als maschinenlesbare Dokumente veröffentlichten Texte möglich. Allerdings befördert der thematische Zusammenhang von *Tausend Tode* in mehrfacher Hinsicht das Auffinden von Verbindungen. Die durch die Wörter und Wortkombinationen der Texte vorgegebenen möglichen Verknüpfungen sind zum einen bei einem gemeinsamen Thema (höchst wahrscheinlich) umfangreicher als bei thematischer Kontingenz. Zum anderen werden die erzeugten Intertexte sowie die vom Rezipienten angesteuerten Kombinationen durch das Grundthema des Korpus beeinflusst. Zwar ist der Suchradius hochgradig individuell und im Rahmen des ‚Wortmaterials‘ von *Tausend Tode* bis zur thematischen Unabhängigkeit erweiterbar. Auch bei (vermeintlich) originärer Suche stellt die Einbettung in das textuelle Umfeld, von dem bei jeder Suche ein Ausschnitt mitgeliefert wird, allerdings häufig eine thematische Anbindung an das Thema ‚Tod‘ her. Die konzeptuelle Einbindung und Maschinenlesbarkeit der Texte ermöglichen es also, dass der

und den Kanonverfechtern mit dem Instrumentarium der quantitativen Literaturwissenschaft zur Seite zu stehen, das in den letzten Jahren unter dem Schlagwort des Distant Reading zu neuer Blüte gelangt ist“, siehe: Gilbert (Anm. 32).

⁶²Die hier vorgenommene genauere Ausdifferenzierung der Wirkung von *Durchschnitt* geht auf eine Diskussion mit Fotis Jannidis und Simone Winko während des Symposiums zurück, denen ich für ihre Anregungen danke. Vgl. detaillierter zu „aufschlussreiche[n] Beobachtungen“ bezüglich der Stimmenvielfalt in *Durchschnitt* auch Annette Gilbert (Anm. 32).

Rezipient zum Mitschreibenden, zum „Textproduzenten“⁶³ avanciert, wobei die Vorgehensweise jener der maschinellen Texterzeugung von *Ox0a* ähnelt. Das „pre-reading“, welches Weel anknüpfend an Giffard als zentrales Moment des Lesens im „docuverse“ beschreibt,⁶⁴ ist bei *Ox0a* dem Publikationsprozess vorgeschaltet. Es vollzieht sich bereits in der Interaktion von menschlichen Produzenten und Programmen. Bei *Tausend Tode* spalten sich dessen Elemente des „finding and selecting“ hingegen zwischen menschlicher Produktion und maschinengestützter Rezeption auf: Die auf diese Weise entstehenden Intertexte sind das Produkt aus der Publikation *Tausend Tode* und der kreativen Einflussnahme des Rezipienten, welches durch die maschinenlesbare Form ermöglicht wird.

Ox0a und *Tausend Tode* legen für das Feld der digitalen Literatur eine Diversifizierung von Lektürepraktiken nahe, welche in Abhängigkeit von den jeweiligen Texten und deren werkförmiger Einbindung verschiedene Aktivitäts- und Aktivierungsmodi des Rezipienten zwischen *Close Reading* und *Distant Reading* bedingen. Im Spannungsfeld von produzenten- und rezipientenseitiger Selektion, Kombination und Korrelation rücken dabei jeweils verschiedene Aspekte des Gesamtwerks in den Fokus.

6 Dynamische Dokumentation

Die räumliche Fixierung an einem in sich unveränderlichen Ort bietet keinen grundlegenden Orientierungspunkt für die Rezeption dynamisch-unabgeschlossener digitaler Werke. Sie wird aber gewissermaßen durch eine zeitliche Strukturierung ersetzt, die allerdings gerade nicht „immutable“, sondern unmittelbar an die spezifischen Abläufe der Produktion, Veröffentlichung und Rezeption geknüpft ist.⁶⁵ Das Moment der Verschränkung dieser Praktiken des

⁶³ Barthes (Anm. 55), 8.

⁶⁴ „Reading in the docuverse thus involves a greater amount of what Giffard calls ‚pre-reading‘. Finding and selecting – the navigation process through the entirety of connected nodes and documents – are intrinsically and continuously part of the actual act of reading.“ Siehe: Adriaan van der Weel, „Feeding our reading machines: From the typographic page to the docuverse“, in: *Digital Studies/Le champ numérique* 6 (2017), <https://www.digitalstudies.org/articles/10.16995/dscn.15/> (letzter Aufruf 4.5.2018).

⁶⁵ Auch Cayley beschreibt den digitalen Text als „complex, temporal object“, bezieht sich hierbei allerdings maßgeblich auf die grundsätzliche Zeitabhängigkeit digitaler Schrift als lesbares Oberflächenphänomen, siehe John Cayley, „Writing on Complex Surfaces“, in: *Dichtung Digital* 7/2 (2005), <http://dichtung-digital.org/2005/2/Cayley/index.htm> (letzter Aufruf 4.5.2018); vgl. ähnlich in Bezug auf das literarische Werk auch Jörgen Schäfer, „Reassembling the Literary. Toward a Theoretical Framework for Literary Communication in Computer-Based Media“, in: Ders./Peter Gendolla (Hg.), *Beyond the Screen: Transformations of Literary Structures, Interfaces and Genres*, Bielefeld 2010. Diese Zeitabhängigkeit bildet die Grundlage der beschriebenen Dynamiken, die hier im Fokus der Betrachtungen stehen.

literarischen Diskurses gewinnt in der „ortlose[n] Dauergegenwart“⁶⁶ digitaler Kommunikation wiederum stabilisierende Funktion: Indem der Rezipient Ausschnitte der ‚textuellen Formationen‘ in einem spezifischen Zustand liest, interpretiert, zitiert, speichert und/oder vervielfältigt, erzeugt er durch seinen Bezug zu diesem Zustand einen zeitlich bestimmbar, metatextuell fixierten Orientierungspunkt, der die unüberschaubaren Textmengen nach bestimmten Kriterien gliedert und sich zwischen die fortlaufenden Erweiterungen schiebt.⁶⁷ Dieser kann als Ausgangs- und Bezugspunkt für die weitere Kommunikation über das Gesamtkorpus sowie für die Beobachtung und Beschreibung von Veränderungen innerhalb desselben dienen. Für die Untersuchung des Gesamtkorpus von *Tausend Tode* oder bereits eines der Texte von *OxOa* könnten sich insbesondere quantitative Analysemethoden als sinnvoll bspw. für die Detektion sprachlicher Charakteristika erweisen, die aufgrund der Stimmenvielfalt der Werke Hinweise entweder auf die sprachliche Gestaltung in Bezug auf bestimmte Themenbereiche oder – aufgrund der Verfügbarkeit der ‚Grundkorpora‘ bei *OxOa* – auf die sprachlichen Konsequenzen der künstlerischen Modellierung geben könnten.⁶⁸

In beiden Beispielen wird die zeitliche Orientierung des Rezipienten auf Ebene der linguistischen Codes produktionsseitig gestützt durch die explizite Angabe von Veröffentlichungsdaten bzw. der zu jeder Version gehörigen Einzeltexte. Jedem nachträglichen Leser ist auf diese Weise die Möglichkeit des unmittelbaren Nachvollzugs zumindest der Grundlage des jeweils festgehaltenen Zustands der ‚Kommunikationssituation‘ zwischen Text und Leser gegeben.⁶⁹ Potenzielle Veränderungen auf Ebene der bibliographischen Codes sind in beiden Fällen

⁶⁶ Stalder (Anm. 1), 147.

⁶⁷ In noch extremerem Maße wird diese ‚Archivierungsfunktion‘ des Rezipienten bei Texten relevant, deren Veränderungen – wie bei *OxOa* durch das dynamisierte Logo angedeutet – eine deutlich höhere Frequenz aufweisen, als dies in den beiden hier diskutierten Beispielen der Fall ist. Dies gilt bspw. für algorithmisch dynamisierte Werke wie Loss Pequeño Glaziers *White faced bromeliads on 20 hectares* (Anm. 59) und für kollaborative Projekte, in denen sich keine klaren Beitragselemente mehr abgrenzen lassen, wie jenes von Davis Douglas initiierte *The World's First Collaborative Sentence* (seit 1994), <http://artport.whitney.org/collection/DouglasDavis/live/Sentence/sentence1.html> (letzter Aufruf 4.5.2018). Analysierbarkeit wird hier jeweils nur noch durch Speicherung von Versionen ermöglicht (vgl. hierzu auch die auf der Eingangsseite des *Sentence* beschriebenen Strategien des Whitney Museums zur ‚preservation of the work‘). Die signifikant unterschiedlichen Formen und Akteure der Textveränderungen verweisen gleichzeitig auf die Notwendigkeit zur weiteren Differenzierung der Relation von Text- und Leseraktivität. Dies gilt insbesondere auch für das Verhältnis der hier verhandelten Phänomene zu den v.a. in den 1990er Jahren theoretisch vielfältig reflektierten Strukturen literarischer Hypertexte (vgl. hierzu exemplarisch Winko (Anm. 36) sowie die Beiträge in George P. Landow (Hg.), *Hyper / Text / Theory*, Baltimore/London 1994).

⁶⁸ Diese Überlegungen zur Nutzung quantitativer Methoden für die Analyse umfangreicher digitaler literarischer Werke basieren auf Gesprächen im Rahmen des Symposiums.

⁶⁹ Veränderungen der Korpora jenseits der Erweiterung durch neue Texte, also etwa durch Änderungen an den Einzeltexten oder durch Wegfallen selbiger, wären nur durch rezipientenseitige ‚Speicherung‘ oder, für *OxOa*, durch Recherche mittels *Wayback Machine* feststellbar.

allerdings ausschließlich durch die rezipientenseitige Speicherung in Form von Screenshots oder die Archivierung der jeweiligen Stylesheets und/oder Lesegeräte sowie der hierfür notwendigen Programme erfassbar.⁷⁰

Während die produzentenseitigen textstabilisierenden Faktoren der Referenzialität und der paratextuellen Rahmung rezipientenseitige Dynamisierungen durch das Stiften eigener Verbindungen ermöglichen, erlangt der Leser im Hinblick auf die destabilisierenden Tendenzen digitaler Literatur – Unabgeschlossenheit, Unüberschaubarkeit und (potenziell nicht gekennzeichnete) Revision – dokumentierende und damit stabilisierende Funktion. In diesem sich durch Interaktion zwischen Produktion bzw. Veröffentlichung und (dokumentierter bzw. dokumentierender) Rezeption konstituierendem dynamischen Netz stellen Vor- und Nachzeitigkeit keine fixen Größen dar. Die wechselseitige Beziehung von Publikations- und Rezeptionszeitpunkten konstituiert aber jeweils zeitlich markierte ‚ Fassungen ‘ der Korpora *und* der Analysen, welche strukturierend wirken und auf diese Weise weitere metatextuelle Referenzen ermöglichen.⁷¹ Während also die Vorstellung einer Textfassung hinsichtlich analoger Literatur ausschließlich am Autor und dessen Produktion orientiert ist (wenngleich auch hier häufig eine [Re-]Konstruktion seitens des Editors erfolgt), lässt sie sich bei digitaler Literatur gerade auf das Verhältnis zwischen Produzenten und Rezipienten beziehen: Letztere haben durch ihre ‚rahmende Intervention‘⁷² entscheidenden Anteil an der Dokumentation – und das heißt wortwörtlich der Zuschreibung eines Dokumentenstatus durch ‚Archivierung‘ – eines Textzustands.⁷³ Diese metatextuellen Fixierungen durch die ‚literarische Öffentlichkeit‘⁷⁴ binden die Texte unter bestimmten Kriterien

⁷⁰Vgl. hierzu Harpold (Anm. 57), 4ff. Dies erscheint für *Ox0a* relevanter als für das genormte Layout von *Tausend Tode* (Frohmann [Anm. 3]). Alle Layout-Ressourcen von *Ox0a* lassen sich auf der Webseite einsehen, Veränderungen auf Ebene der linguistischen Codes können z.B. in der Rubrik „Aktuelles“ über die *Wayback Machine* nachvollzogen werden.

⁷¹Der Frohmann-Verlag und *Ox0a* explizieren diese wechselseitige Beziehung, indem Teile der jeweiligen Netze in Form von Rezensionen und Kommentaren auf der Webseite des Verlags, auf der *Tausend Tode* (Frohmann [Anm. 3]) angeboten wird, bzw. auf den zu den Einzeltexten gehörigen Unterwebseiten von *Ox0a* verlinkt sind.

⁷²Vgl. Drucker, *What is?*, 58.

⁷³In diesem Sinne verweist auch Kirschenbaum darauf, dass „[i]n the digital realm there is a real sense [...] in which archive can only ever be a verb, marking the latent potential for reconstruction and reconstitution“, siehe Matthew Kirschenbaum, „The .xtual Condition. Digital Humanities, Born-Digital Archives, and the Future Literary“, in: *Digital Humanities Quarterly* 7/1 [2013], Abschn. 12, <http://www.digitalhumanities.org/dhq/vol/7/1/000151/000151.html> (letzter Aufruf 4.5.2018). Vgl. zu dieser dokumentierenden Funktion des Rezipienten auch Kirschenbaums Beschreibung der ‚Überlieferungsgeschichte‘ von William Gibsons und Dennis Ashbaughs *Agrippa (A Book of the Dead)* (1992), <http://agrippa.english.ucsb.edu> (letzter Aufruf 4.5.2018), „which was intended to disappear from sight, [but] yet is one of the most stable an accessible electronic objects I know“ (Matthew Kirschenbaum, *Mechanisms. New Media and the Forensic Imagination*, Cambridge, MA, 2008, XII).

⁷⁴Reuß (Anm. 42), 10.

wiederum zu Referenz-Netzwerken zusammen, deren Ausgestaltung von der jeweiligen textuellen Grundlage ebenso abhängig ist wie von den spezifischen heuristischen Schwerpunktsetzungen der Interpreten. Die literarische Wirkungsgeschichte dokumentiert sich mithin prozesshaft in Interaktion mit der fortgesetzten Produktion und Publikation der Werke selbst. Werk und Wirkungsgeschichte bringen sich daher tatsächlich „wechselseitig allererst im Verlauf der Geschichte hervor“.⁷⁵

7 Interferenzen von digitalen und analogen Ordnungsmustern

Bisher standen maßgeblich die durch digitale Literatur ausgelösten Verschiebungen, also die Differenzen zu analogen Ordnungsmustern, im Zentrum der Betrachtungen. Allerdings sind diese Differenzen nicht im Sinne unüberbrückbarer Gegensätze zu verstehen. Vielmehr zeichnen sich sowohl *Tausend Tode* als auch *Ox0a* durch Wechselwirkungen zwischen digitalen und analogen Ordnungsmustern aus. Dies erzeugt „complex feedback loops [that] connect humans and machines, old technologies and new, language and code, analog processes and digital fragmentations“.⁷⁶ Dieser Umstand soll abschließend anhand einzelner technologischer und institutioneller Spezifika zumindest angedeutet werden.

Die beiden hier untersuchten Beispiele weisen signifikante Unterschiede hinsichtlich der beobachtbaren Interferenzen von digitalen und analogen Paradigmen auf: Digitale Paradigmen realisieren sich bei *Ox0a* maßgeblich im Rahmen der Produktion der Texte (Algorithmen, Sampling digital verfügbarer Texte, Referenz auf digitale Kommunikationskanäle wie Facebook, Wikipedia), der Präsentation des Kollektivs (Webseite, Oszillieren zwischen Gruppen- und Werktitel) sowie in der Open-Access-Politik hinsichtlich des Zugangs zu den Texten. Diese Hinwendung zu digitalen Paradigmen wird allerdings dadurch konterkariert, dass die Einzeltext-Publikationen in den allermeisten Fällen als PDFs erfolgen,⁷⁷ also in einem Format, welches in hohem Maße durch die Strukturierungspraktiken der Gutenberg-Ära geprägt ist⁷⁸ und bereits in seiner Benennung auf Latours „immutable mobiles“ zu verweisen scheint. Diese, auch durch die Angabe der

⁷⁵ So ebd. – allerdings in Bezug auf die der Publikation nachfolgende Wirkungsgeschichte.

⁷⁶ So Hayles (Anm. 18), 31, hinsichtlich des von ihr unter dem Terminus der ‚intermediation‘ verhandelten Phänomens der „dialectic between analog and digital representations“ (ebd., 6).

⁷⁷ Zusätzlich wird meist eine Kaufoption angeboten, die entweder auf Print-on-Demand-Seiten oder auf Amazon verlinkt ist. Die einzigen Texte, welche unmittelbar auf der Webseite erscheinen, sind Hannes Bajohrs *Wendekorpus* (2014), <http://0x0a.li/de/text/wendekorpus/>, und dessen *Identitätsdrama* (2014), <http://0x0a.li/de/text/identitaetsdrama/> (letzter Aufruf der Links 4.5.2018).

⁷⁸ Vgl. Gitelman (Anm. 43), 118f. und 131.

Veröffentlichungsdaten gestützte ‚Dokumentation‘ knüpft an analoge Praktiken der Stabilisierung und Institutionalisierung an und steht im Kontrast zur Dynamik digitaler Strukturen.⁷⁹ Diese Praktik kann als Reverenz an die grundlegenden Mechanismen des literarischen Diskurses verstanden werden, für dessen Wechselverhältnis von Text und interpretierender Lektüre ‚Beobachtbarkeit‘ unabdingbar ist, die nur im Rahmen einer gewissen Stabilität erreicht werden kann.⁸⁰

Die Texte in *Tausend Tode* hingegen sind nach analogen Prinzipien produziert und die Distribution über den Verlag bzw. angeschlossene kommerzielle Dienste folgt den klassischen Prinzipien des Buchmarkts.⁸¹ Die Verschränkung von Produktion, Publikation und Rezeption sowie die stetigen Wechsel von Genres und Autoren gestalten aber insbesondere das Lektüreerlebnis als ein spezifisch digitales. Die typographische Gestaltung der ‚Ausgabe‘ erfolgt – ähnlich wie für *Ox0a* festgestellt – im Spannungsfeld von digitalen dynamisierenden und analogen schriftfixierenden Codierungen, indem bestimmte Zeilenumbrüche innerhalb des Korpus im Gegensatz zum grundsätzlich dynamisierten Erscheinungsbild der Schrift stabilisiert sind. Die auf diese Weise ‚markierten‘ Textstellen erscheinen dadurch in Bezug auf ihr dynamisiertes Umfeld besonders betont.

Inwieweit bezüglich der Interferenzen von analogen und digitalen Ordnungsmustern von einer „period of ‚imitation““⁸² gesprochen werden kann, die überwunden werden wird, bleibt abzuwarten. Dies mag für die expliziten Abgrenzungsbewegungen digitaler Literatur von analogen Paradigmen bzw. die selbstreflexiven Verweise auf den eigenen digitalen Status gelten, wie sie insbesondere bei *Ox0a* feststellbar sind. In Bezug auf bestimmte Phänomene wie z.B. die Dokumentation von Versionen und die Schriftgestaltung verweisen *Tausend Tode* und *Ox0a* allerdings auf die spezifische bedeutungstragende Funktion analoger Paradigmen für den literarischen Diskurs und/oder die Textsemantik. In diesem Sinne lässt sich ebenso postulieren, dass sich digitale Literatur weiterhin spezifischer analoger Signifikationsstrategien und Publikationspraktiken bedienen wird. Diese Übernahmen verändern sowohl das ‚aufnehmende‘ mediale Format als auch die Wirkung der angeeigneten Strategien und Praktiken.

⁷⁹Zum PDF als „archival standard“ vgl. ebd., 116.

⁸⁰Von einer derartigen Reverenz zeugen auch die Publikationen der *Electronic Literature Organization*, welche zwar im *Open Access* erscheinen, allerdings mit dem expliziten Hinweis, dass weder kommerzielle Nutzung noch Modifikation erlaubt sind; vgl. <http://collection.eliterature.org/1/aux/about.html> (letzter Aufruf 4.5.2018).

⁸¹Differenzierungen entlang einer Grenze von Amateurstatus und Profession erscheinen in diesem Zusammenhang aktuell zwar naheliegend, erweisen sich aber als problematisch, da sich die Distributionspraktiken von digitaler Literatur in einem frühen Entwicklungsstadium befinden, dessen weitere Ausgestaltung noch nicht absehbar ist.

⁸²So van der Weel maßgeblich in Bezug auf „the typographical habit“, siehe ders. (Anm. 64).

8 Dynamisierung und Restabilisierung als Paradigmen digitaler Textualität

Die beiden untersuchten Beispiele genuin digitaler Literatur erweisen sich in ihren Produktions- und Distributionspraktiken und vor allem hinsichtlich der Paradigmen der Ausgestaltung literarischer Zusammenhänge von den Merkmalen der Kultur der Digitalität affiziert: Algorithmizität, neue Formen der Gemeinschaftlichkeit und insbesondere das Prinzip der Referenzialität⁸³ bestimmen mit unterschiedlichen Schwerpunktsetzungen die ‚Funktionsweise‘ und Wirkung der Korpora und beeinflussen – so wurde hier postuliert – auf diese Weise auch das Rezeptionsverhalten.

Zentrale Strukturen, welche die Konzeptionierung literarischer Textualität prägen, unterliegen in diesem Zusammenhang signifikanten Veränderungen. Dies gilt insbesondere für das (ohnehin komplizierte) Beziehungsgefüge zwischen Autorschaft, Text und Werk sowie für den Prozess zwischen Textproduktion und Lektüre. Dabei ist weniger von einem Bedeutungsverlust einzelner Parameter und Instanzen als von wechselwirkenden Verschiebungen ihrer Funktionen zu sprechen. Diese rühren maßgeblich daher, dass der einzelne Text direkt als Teil eines ideellen und materiellen Referenznetzes publiziert wird. Dieses stabilisiert die Beiträge als Texte und verleiht den Korpora durch Strategien der Formation und Präsentation Werkstatus, dem jedoch der Absolutheitsanspruch analoger Werke fehlt: Die referenzielle Einbindung folgt zwar spezifischen Kriterien, diese sind aber gerade *nicht* verbindlich. Die Texte können genauso oder in veränderter Form, insgesamt oder in Ausschnitten ebenso in anderen Formationen erscheinen. Gleichzeitig unterliegen die Werke durch das Hinzukommen weiterer Texte permanenten Dynamisierungen auf materieller und semantischer Ebene. Beziehungsstiftende Zuschreibungen, die sich an Autorschaft oder Genre⁸⁴ orientieren, erscheinen für die untersuchten Beispiele daher nicht nur unmöglich, sondern erweisen sich als unnötig, weil der einzelne Text unmittelbar in seinem

⁸³Diese drei „Formen der Digitalität“ sind für Stalder kennzeichnend für die „neue, spezifische kulturelle Umwelt, die vor unseren Augen deutliche Gestalt annimmt und mehr und mehr andere kulturelle Konstellationen überlagert beziehungsweise an den Rand drängt“, siehe Stalder (Anm. 1), 95.

⁸⁴Dies gilt hier für die Zuschreibung der Einzeltexte zu einem etablierten literarischen Genre, wohingegen die systematische Ausdifferenzierung von ‚Gattungen‘ der digitalen Literatur trotz einiger Ansätze bislang ein Desiderat darstellt. Die vergleichende Betrachtung zweier auf verschiedenen Ebenen divergierender, dabei aber gleichzeitig übereinstimmender Funktions- und Wirkweisen zeitigender Beispiele sowie die punktuell vorgenommenen Vergleiche mit anderen Formen digitaler Literatur verweisen allerdings bereits darauf, dass eine Differenzierung digitaler literarischer Genres anhand ihrer Strukturen wahrscheinlich nicht für gesamte Texte funktioniert. Nicht zuletzt die Vervielfältigung von Schreibpraktiken und Schriftebenen lassen Differenzierungen für verschiedene Parameter gewinnbringender erscheinen, die auf unterschiedlichen Ebenen Gemeinsamkeiten und Unterschiede zwischen digitalen literarischen Texten systematisierbar machen.

jeweiligen Referenzfeld rezipiert werden kann. Die Gestaltung der referenziellen Strukturen sowie der sie dynamisch konstituierenden Texte rücken das Verhältnis des Einzeltextes zu seinem jeweils spezifischen Umfeld sowie vergleichende Betrachtungen der Wirkung des Textes in anderen oder sich verändernden Kompositionen in den Fokus. Die anhand der beiden Beispiele beobachtbaren Format- und Formatierungsunterschiede verweisen in diesem Zusammenhang auf die prinzipielle Notwendigkeit der Differenzierung verschiedener Ausgabeformen von digitaler Literatur jenseits verallgemeinernder Abgrenzungen *des* digitalen Mediums von analogen Repräsentationsformen. Jene Ausgabeformen prägen, vergleichbar mit der unterschiedlichen Formatierung analoger Literatur etwa durch die ‚Ausgabe‘ in einem Buch oder einer Zeitschrift, die grundsätzlichen Erscheinungsweisen der Texte. Sie müssen daher aufgrund ihrer formierenden Wirkungen auf die in ihnen verkörperten Texte in eine Untersuchung der jeweiligen Text-Kontext-Verhältnisse einbezogen werden.

Für einen an digitalen Paradigmen orientierten Textbegriff erweist sich auf Grundlage der untersuchten Beispiele insbesondere das für die Korpora als entscheidend herausgearbeitete Wechselverhältnis von dynamisierender Destabilisierung und formierender Restabilisierung als konstitutiv: Textualität realisiert und dokumentiert sich in der Verschränkung von Produktion, Publikation und Rezeption als bedeutungstiftende Relation zwischen markierten Einzel-elementen und einem Werk als im Vollzug beobachtbares Projekt mit vielen Akteuren sowie einer spezifisch organisierten medialen Verkörperung.⁸⁵

Literatur

Sämtliche digitalen Referenzen wurden letztmalig am 4.5.2018 eingesehen.

ANd-OR, *dadaOverload* (2016) <http://www.and-or.ch/dadaoverload/>.

Bajohr, Hannes, „Das Reskilling der Literatur. Einleitung zu Code und Konzept“, in: Ders. (Hg.), *Code und Konzept. Literatur und das Digitale*, Berlin 2016, 7–21.

Bajohr, Hannes, *Durchschnitt* (2015), <http://0x0a.li/de/text/durchschnitt/>.

Bajohr, Hannes, *Identitätsdrama* (2014), <http://0x0a.li/de/text/identitaetsdrama/>.

Bajohr, Hannes, *Wendekorpus* (2014), <http://0x0a.li/de/text/wendekorpus/>.

Bajohr, Hannes/Weichbrodt, Gregor, *Glaube Liebe Hoffnung* (2015), <http://0x0a.li/de/text/glaube-liebe-hoffnung/>.

Bajohr, Hannes/Weichbrodt, Gregor, *These*, <http://0x0a.li/de/these/>.

Barthes, Roland, „Vom Werk zum Text“ [1971], in: Stephan Kammer/Roger Lüdeke (Hg.), *Texte zur Theorie des Textes*, Stuttgart 2005, 40–51.

Barthes, Roland, *S/Z*, Frankfurt a. M. 1976.

Cayley, John, „The Code is not the Text (unless it is the Text)“, in: *electronic book review* (10. September 2002), <http://www.electronicbookreview.com/thread/electropoetics/literal>.

Cayley, John, „Writing on Complex Surfaces“, in: *Dichtung Digital 7|2* (2005) <http://dichtung-digital.org/2005/2/Cayley/index.htm>.

⁸⁵ Ich danke Thomas Ernst für seinen Kommentar sowie Maria Hinzmann und Sophia Krebs für ihre kritische Lektüre und hilfreichen Hinweise zu diesem Beitrag.

- Davis, Douglas (Initiator), *The World's First Collaborative Sentence* (seit 1994), <http://artport.whitney.org/collection/DouglasDavis/live/Sentence/sentence1.html>.
- Drucker, Johanna, *What is? (a letter, writing, a word's body, a document, graphic textuality ...)* *Nine epistemological essays*, Victoria, TX, 2013.
- Foucault, Michel, „Was ist ein Autor?“, in: Ders., *Schriften zur Literatur*, hg. von Daniel Defert/François Ewald, Frankfurt a. M. 2003 [1974], 234–270.
- Frohmann, Christiane (Hg.), *Tausend Tode schreiben*, Version 3.1., Berlin 2015.
- Gibson, William/Ashbaugh, Dennis, *Agrippa (A Book of the Dead)* (1992), vgl. <http://agrippa.english.ucsb.edu>.
- Gilbert, Annette, „Möglichkeiten von Text im Digitalen“. Ästhetische Urbarmachung von korpuslinguistischen Analysetools in der generativen Literatur der Gegenwart am Beispiel des Textkollektivs 0x0a“, in: *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* 91/203 (2017), DOI <https://doi.org/10.1007/s41245-017-0038-y>.
- Gitelman, Lisa, *Paper Knowledge. Toward a Media History of Documents*, Durham/London 2014.
- Glazier, Loss Pequeño Glazier, *White faced bromeliads on 20 hectares* (1999), http://collection.eliterature.org/1/works/glazier__white-faced_bromeliads_on_20_hectares.html.
- Grigat, Guido (Initiator), *23:40* (seit 1997), <http://www.dreiuundzwanzigvierzig.de>.
- Harpold, Terry, *Ex-foliations. Reading Machines and the Upgrade Path*, Minneapolis 2009.
- Hartling, Florian, *Der digitale Autor. Autorschaft im Zeitalter des Internets*, Bielefeld 2009.
- Hasecke, Jan Ulrich (Initiator), *Das Generationenprojekt* (seit 1997), <https://www.generationenprojekt.de>.
- Hayles, Nancy K., *My Mother was a Computer. Digital Subjects and Literary Texts*, Chicago 2005.
- Ivanovic, Christine, „Die Vernetzung des Textes. Im Möglichkeitsraum digitaler Literaturanalyse“, in: *Zeitschrift für digitale Geisteswissenschaften* (2017), Kap. 4, o. S., DOI https://doi.org/10.17175/2016_010.
- Julia Nantke, „Multiple Autorschaft als digitales Paradigma und dessen Auswirkungen auf den Werkbegriff“, in: Svetlana Efimova (Hg.), *Textpraxis. Digitales Journal für Philologie 2* (2018), Sonderausgabe #3, DOI <https://doi.org/10.17879/77159516645>.
- Kirschenbaum, Matthew, „The .txtual Condition. Digital Humanities, Born-Digital Archives, and the Future Literary“, in: *Digital Humanities Quarterly* 7/1 (2013), <http://www.digitalhumanities.org/dhq/vol/7/1/000151/000151.html>.
- Kirschenbaum, Matthew, *Mechanisms. New Media and the Forensic Imagination*, Cambridge, MA, 2008.
- Klinger, Claudia (Initiatorin), *Beim Bäcker* (1996–2000).
- Kraus, Dagmar, *LENZ* (2016), <http://0x0a.li/de/text/lenz/>.
- Landow, George P. (Hg.), *Hyper / Text / Theory*, Baltimore/London 1994.
- Landow, George P., *Hypertext 3.0. Critical Theory and New Media in an Era of Globalization*, Baltimore 2006.
- Latour, Bruno, „Drawing Things Together“, in: Michael Lynch/Steve Woolgar (Hg.), *Representation in Scientific Practice*, Cambridge, MA, 1990, 19–68.
- Lotman, Juri M., *Kunst als Sprache. Untersuchungen zum Zeichencharakter von Literatur und Kunst*, Leipzig 1981.
- Martínez, Matías, „Autorschaft und Intertextualität“, in: Ders./Fotis Jannidis/Gerhard Lauer u.a. (Hg.), *Die Rückkehr des Autors. Zur Erneuerung eines umstrittenen Begriffs*, Tübingen 1999, 465–479.
- McGann, Jerome J., *The Textual Condition*, Princeton 1991.
- Nelke, Roger (Initiator), *Magische Welt. İja Macár* (1999–2008), http://www.drachental.de/ijamacar/im_index.htm.
- Penguin Verlag (Initiator), *A million penguins* (2007).
- Reuß, Roland, „Text, Entwurf, Werk“, in: *Text. Kritische Beiträge* 10 (2005), 1–12.

- Schäfer, Jörgen, „Reassembling the Literary. Toward a Theoretical Framework for Literary Communication in Computer-Based Media“, in: Ders./Peter Gendolla (Hg.), *Beyond the Screen: Transformations of Literary Structures, Interfaces and Genres*, Bielefeld 2010, 25–70.
- Schmitt, Stefan, „Mehr Aber als Ja“, in: *DIE ZEIT* 44 (29. Oktober 2015), 1.
- Simanowski, Roberto, „Autorschaft und digitale Medien. Eine unvollständige Phänomenologie“, in: Lucas Marco Gisi/Urs Meyer/Reto Sorg (Hg.), *Medien der Autorschaft. Formen literarischer (Selbst-)Inszenierung von Brief und Tagebuch bis Interview und Fotografie*, München 2013, 247–262.
- Simanowski, Roberto, *Interfictions. Vom Schreiben im Netz*, Frankfurt a. M. 2002.
- Spoerhase, Carlos, „Was ist ein Werk? Über philologische Werkfunktionen“, in: *Scientia Poetica* 11 (2007), 276–344.
- Stalder, Felix, *Kultur der Digitalität*, Frankfurt a. M. 2016.
- Tabbi, Joseph, „Locating the Literary in New Media“, in: *electronic book review* (23. September 2008), <http://www.electronicbookreview.com/thread/criticalecologies/interpretive>.
- Ulshöfer, Susann (Initiatorin), *Die Perlen von Caala-Elen* (1999–2008), <https://web.archive.org/web/20070314051902/http://www.zauberfee.de:80/zauberbuch/caala-elen/index.htm>.
- van der Weel, Adriaan, „Feeding our reading machines. From the typographic page to the docuverse“, in: *Digital Studies/Le champ numérique* 6 (2017), <https://www.digitalstudies.org/articles/https://doi.org/10.16995/dscn.15/>.
- Weichbrodt, Gregor, *Chicken Infinite* (2014), <http://0x0a.li/de/text/chicken-infinite/>.
- Weichbrodt, Gregor, *Dictionary of non-notable Artists* (2016), <http://0x0a.li/de/text/dictionary-of-non-notable-artists/>.
- Winko, Simone, „Lost in Hypertext? Autorkonzepte und neue Medien“, in: Dies./Fotis Jannidis/Gerhard Lauer u.a. (Hg.), *Die Rückkehr des Autors. Zur Erneuerung eines umstrittenen Begriffs*, Tübingen 1999, 511–533.
- Zimmermann, Heiko, *Autorschaft und digitale Literatur. Geschichte, Medienpraxis und Theoriebildung*, Trier 2015.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Wie und wozu analysiert man Blogs? Überlegungen zwischen *Digital Philology*, *Distant Reading* und Kulturpoetik

Jörg Schuster

1 Wo bleibt die elektronische Avantgardeliteratur?

Das Verhältnis von Literatur und elektronischen Medien hat sich anders entwickelt, als dies in den 1990er Jahren – als Computer und Internet zu Alltagsphänomenen avancierten – erwartet worden war. Der Hypertext schien damals die Möglichkeit einer neuen experimentellen Avantgarde zu eröffnen: Erzähltechnische Innovationen der Moderne wie multiperspektivisches und nicht-lineares Erzählen wurden im Zuge der neuen elektronischen Möglichkeiten radikalisiert, grundlegende literaturtheoretische Begriffe wie ‚Text‘ und ‚Autor-schaft‘ schienen neu definiert werden zu müssen.¹ Als neue experimentelle Form hat sich der Hypertext jedoch nicht durchgesetzt. Die Gründe dafür sind vielfältig: Für die Autorinnen und Autoren erwies sich die Programmierung als extrem aufwendig, es fehlte ein breiter Absatzmarkt für die neuartigen Produkte, und ohnehin waren die letzten Jahrzehnte nicht gerade eine Blütezeit des literarischen Experiments – angesagt waren, gerade vor dem Hintergrund radikaler technisch-medialer und politisch-sozialer Veränderungen wie der digitalen Revolution und der Globalisierung, ein eher traditioneller populärer Realismus und das Ideal der Leserfreundlichkeit. Während etwa im Bereich der Populärkultur Computerspiele aufgrund kontinuierlicher Innovationen boomten, hat sich die Hoffnung auf eine neue literarische Avantgarde im Zeichen des elektronischen Experiments nicht

¹Vgl. Ruth Nestvold, „Das Ende des Buches. Hypertext und seine Auswirkungen auf die Literatur“, in: Martin Klepper/Ruth Mayer/Ernst-Peter Schneck (Hg.), *Hyperkultur. Zur Fiktion des Computerzeitalters*. Berlin/New York 1996, 14–30; Roberto Simanowski, *Interfictions. Vom Schreiben im Netz*, Frankfurt a.M. 2002.

J. Schuster (✉)
Universität Frankfurt, Frankfurt am Main, Deutschland
E-Mail: schuster@lingua.uni-frankfurt.de

erfüllt. Die mediale Revolution zog keine literarische Revolution nach sich, jedenfalls bisher noch nicht.

Welche Folgen hatte das neue Medium aber dann für das gegenwärtige literarische Feld? Überblickt man das Untersuchungsgebiet ‚Literatur und Internet‘, so denkt man heute eher an Online-Buchhandel über Amazon, mehr oder weniger unseriöse literarische Online-Datenbanken wie das *Projekt Gutenberg-DE*² oder, von der Literaturwissenschaft in letzter Zeit mit besonderer Vorliebe herangezogen, an Autorschaftsinszenierungen von Schriftsteller*innen im Netz, etwa in Form von Homepages oder Facebook-Auftritten, die der Pflege der eigenen Leser*innen-Fangemeinde dienen.³ Dass das Netz zum Aufbegehren gegen das etablierte Literatursystem und die Kommerzialisierung genutzt wurde, blieb die Ausnahme – so im Fall der gegen das ‚Machtmittel Buch und Buchbetrieb‘⁴ gerichteten Provokation, die Elfriede Jelinek mit der kostenlosen Veröffentlichung ihres ‚Privatromans‘ *Neid* im Internet anstrebte. Von solchen illustren Ausnahmen abgesehen, scheint das Internet – auf den ersten Blick – bislang eher den traditionellen Buchmarkt und herkömmliche literarische Genres zu stabilisieren als radikal Neues hervorzubringen.

2 Populärkultur – Kundenrezensionen, Fan-Fictions, Literaturplattformen, Blogs

Die Einschätzung, dass es im Netz nichts Neues gebe, hat allerdings viel mit unserem literaturwissenschaftlichen Blick zu tun, der traditionell elitär auf ‚seriöse‘ Autor*innen, auf ‚Höhenkammliteratur‘ zentriert ist. Das Internet, insbesondere das Web 2.0, steht hingegen für Massenkommunikation in dem Sinne, dass jede*r Nutzer*in nicht nur Rezipient*in ist, sondern zugleich auch Produzent*in, Akteur*in werden kann. Das hat weitreichende Folgen: Herkömmliche Grenzen wie die zwischen Privatem und Öffentlichem sowie Professionalität und Nicht-Professionalität lassen sich nicht mehr genau markieren. Hinzu kommt, dass die Aufmerksamkeit in weitaus stärkerem Maß als in traditionellen Medien durch rein quantitative Kriterien gelenkt wird: Ob nun Facebook oder YouTube – was zählt, ist, innerhalb möglichst kurzer Zeit möglichst viele Klicks zu erreichen oder möglichst oft ‚geliked‘ zu werden. Auf diese Weise kann man ‚über Nacht‘ zum Star werden und genauso schnell wieder in Vergessenheit geraten. Benannte Charles Baudelaire bereits 1863 in seinem Essay *Le peintre de la vie moderne* bekanntlich die sich

² <https://gutenberg.spiegel.de/> (letzter Aufruf 24.9.2017).

³ Vgl. exemplarisch die entsprechenden Beiträge in: Christine Künzel/Jörg Schönert (Hg.), *Autorinszenierungen. Autorschaft und literarisches Werk im Kontext der Medien*, Würzburg 2007; Christoph Jürgensen/Gerhard Kaiser (Hg.), *Schriftstellerische Inszenierungspraktiken. Typologie und Geschichte*, Heidelberg 2011.

⁴ Elfriede Jelinek, *Keine Anweisung. Keine Auszahlung, kein Betrag, kein Betrug*, <http://www.elfriedejelinek.com/fanmerk.htm> (letzter Aufruf 24.9.2017).

durch „das Vergängliche, das Flüchtige, das Zufällige“⁵ auszeichnende *Mode* als ein entscheidendes Kriterium für Modernität, so erscheint diese Kurzlebigkeit im Internet-Zeitalter ins Extrem gesteigert. Die Frage ist, ob man diese Entwicklung bedenklich finden oder ob man ihr nicht vielmehr produktive Potenziale gerade im Bereich des Ästhetischen attestieren sollte – genau wie Baudelaire vor mehr als 150 Jahren das Flüchtige der Mode – neben dem Ewigen und Unwandelbaren – als „die eine Hälfte“⁶ des Schönen und der Kunst definiert hatte.

Dieser Hälfte des Schönen kann sich die Literaturwissenschaft nicht entziehen. Sie sollte daher nicht ignorieren, dass sich auch im Bereich der literarischen Kommunikation neue Formen und Institutionen herausgebildet haben. Es gilt, den Blick zu lenken auf populärkulturelle Phänomene wie Kundenrezensionen, Fan-Fictions, Mitschreibprojekte im Rahmen von Literaturplattformen sowie auf Blogs, die ich in den Mittelpunkt meiner Überlegungen stellen möchte (und daher zunächst einmal übergehe). Man kann über die Masse und Ubiquität der – oft dilettantischen – Literaturkritik und literarischen Produktion im Netz also die Nase rümpfen; vor allem aber kann man sie aufgrund ihrer einfachen Verfügbarkeit im digitalen Speichermedium *analysieren* – und darin besteht ein entscheidender Vorteil für den/die Literaturwissenschaftler*in. Noch nie ließen sich wohl Prozesse literarischer Kommunikation so flächendeckend und leicht untersuchen wie heute. blieb etwa die Beschäftigung mit der Rezeptionsgeschichte im Hinblick auf vergangene Epochen aufgrund einer meist geringen Datenmenge oft weitgehend spekulativ, so sind aktuellen Untersuchungen kaum Grenzen gesetzt – außer der Scheu vor den vermeintlichen Niederungen der Alltagskultur.

Im Bereich der Literaturkritik führt die sich im Zeichen des Digitalen vollziehende Dezentralisierung (die nicht über die zunehmende ökonomische Zentralisierung hinwegtäuschen sollte) in verstärktem Maße jene „dialogischen und populären Traditionen“⁷ fort, die sie seit ihrem Beginn in der Epoche der Aufklärung besaß. Auch Literatur-Plattformen bieten – je nachdem, in welchem Maß sie nicht-hierarchische Kommunikation erlauben – die Möglichkeit dezentraler Interaktion im Sinne digitaler Text-Werkstätten.⁸ Zunehmend an Popularität

⁵ Charles Baudelaire, „Der Maler des modernen Lebens“, in: Ders., *Sämtliche Werke und Briefe*, Bd. 5: *Aufsätze zur Literatur und Kunst 1857–1860*, hg. von Friedhelm Kemp/Claude Pichois, München/Wien 1989, 226.

⁶ Baudelaire (Anm. 5).

⁷ Thomas Anz, „Literaturkritik in Zeiten des Internets. Thesen zu ihrer Verteidigung und Bedenken“, in: *literaturkritik.de*, 12. Dezember 2010, http://www.literaturkritik.de/public/rezension.php?rez_id=15120 (letzter Aufruf 24.9.2017); vgl. ferner Thomas Wegmann, „Warentest und Selbstmanagement. Literaturkritik im Web 2.0 als Teil nachbürgerlicher Wissens- und Beurteilungskulturen“, in: Matthias Beilein/Claudia Stockinger/Simone Winko (Hg.), *Kanon, Wertung und Vermittlung. Literatur in der Wissensgesellschaft*, Berlin/Boston 2012, 279–291.

⁸ Vgl. Jörg Schuster, „Kreativer Austausch oder digitaler Masochismus? Internet-Schreibplattformen beleben die literarische Kommunikation, wenn sie die Chancen der Interaktivität und Dezentralisation nutzen – die Beispiele *literaturcafe.de* und *leselupe.de*“, in: *literaturkritik.de*, 4. April 2013, <https://literaturkritik.de/id/17760> (letzter Aufruf 24.9.2017).

gewann in den letzten Jahren die Textsorte ‚Fan-Fiction‘, deren pädagogisch-kreativen Wert für den handlungs- und produktionsorientierten Unterricht inzwischen auch die Literaturdidaktik erkannt hat. Auch jenseits indizierender Einzelfälle wie dem aus einer Fan-Fiction (zu *Twilight*) hervorgegangenen erotischen Bestseller-Roman *Fifty Shades of Grey* liegt die kulturhistorische und kulturdiagnostische Bedeutung des Phänomens auf der Hand. Es dürfte kaum einen geeigneteren Indikator nicht allein für den quantitativen Erfolg, sondern genauer für die „soziale Energie“ (Stephen Greenblatt)⁹ literarischer Werke geben als diese Form der kreativen Rezeption.¹⁰

Das Weiter- und Umschreiben von Handlungs- und Figurenmustern als idealtypische Form von Intertextualität, in deren Rahmen häufig die Geschlechtsidentitäten und die sexuelle Orientierung von Figuren transformiert werden, ist allein für die *Gender Studies* und *Queer Studies* von unschätzbarem Wert. Generell ist hier auf der Basis einer enormen Datenfülle zu beobachten, wie *Verständigung über das kulturelle Wissen einer Zeit* stattfindet, wie popkulturell oder, wenn man so will, trivial dieses Wissen auch immer sein mag. Der individuelle kreative Umgang mit Prätexten ist dabei aus literatur- und kulturwissenschaftlicher Sicht ebenso aufschlussreich wie die Anschlusskommunikation zwischen den User*innen; in unzähligen Kommentaren zu den auf den Plattformen eingestellten Texten wird darüber diskutiert, wie plausibel die jeweilige Transformation des literarischen Originals ist, inwiefern also die individuelle Abweichung von einem adäquaten Verständnis des Prätexts zeugt. Nachvollziehbar werden auf diese Weise die Regeln der Selbstverständigung innerhalb einer literarisch interessierten Community. Im Bereich der Fan-Fiction präsentiert sich dem/der Literaturwissenschaftler*in somit ein umfassendes Archiv kulturellen Wissens.

Wie aber gehen wir als Literaturwissenschaftler*in mit solchen monströsen Archiven um, wie nutzen wir die enorme Menge an Daten, die populärkulturelle Phänomene wie Kundenrezensionen, Literaturplattformen, Fan-Fictions und Blogs bereitstellen? Und welche Konsequenzen sind für die literaturwissenschaftliche Methodologie zu ziehen?

3 Methodologische Perspektiven: *Digital Philology*, *Distant Reading*, Kulturpoetik

Die Tatsache, dass – in Bezug auf das gegenwärtige literarische Feld wie auch retrospektiv im Hinblick auf vergangene literaturhistorische Epochen – eine immer größere Masse elektronischer Daten verfügbar ist, verändert die Literaturwissenschaft nachhaltig, sie eröffnet das Feld der *Digital Philology*. Über die Art und

⁹Stephen Greenblatt, *Verhandlungen mit Shakespeare. Innenansichten der englischen Renaissance*, übers. David Cackett, Berlin 1990, 24.

¹⁰Auf der deutschsprachigen Seite FanFiktion.de finden sich so derzeit allein mehr als 40.000 Texte zu *Harry Potter*, auf der englischsprachigen fanfiction.net ist es rund die zehnfache Menge.

Weise, in der sie das tut, über die Frage, welche methodologischen Perspektiven sich also im digitalen Zeitalter ergeben, muss dringend diskutiert werden. Bekanntlich hat Franco Moretti für den literaturwissenschaftlichen Umgang mit großen Textkorpora den prägnanten Begriff *Distant Reading* geprägt.¹¹ Während die Methode des *Close Reading* seit den 1950er und 1960er Jahren versuchte, wenige singuläre, kanonische literarische Texte möglichst genau zu analysieren, geht es ihm um eine programmatische Hinwendung „vom Außergewöhnlichen zum Alltäglichen, von den Einzel- zu den Massenerscheinungen“.¹² Was ihm vorschwebt, ist die Analyse jener 99 % der literarischen Texte, die zugunsten des 1 % kanonisierter Texte vergessen wurden. Das Ziel ist nicht mehr und nicht weniger als eine neue Literaturgeschichte im Sinne einer *Comparative Morphology*, die herkömmliche Epochenkonstruktionen infrage zu stellen vermag. Moretti geht es um die Untersuchung kleinster Einheiten wie einzelner Motive oder Erzähltechniken an einer möglichst großen Zahl von Texten, um etwa die historische Transformation literarischer Genres beschreiben zu können. Ziel ist es, die sich verändernde Häufigkeit des Auftretens dieser Einheiten festzustellen und neue Eigenschaften von Texten ans Licht zu bringen: „Die quantitative Herangehensweise fördert die Probleme ans Tageslicht, der Weg über die Form führt zu ihrer Lösung.“¹³ Praktiziert wird somit ein „formalism without close reading“.¹⁴

Allerdings sieht sich Moretti, wie er in seinem Aufsatz *The Slaughterhouse of Literature*¹⁵ eingesteht, mit einem gravierenden Problem konfrontiert: Die Literaturgeschichte soll umgeschrieben werden, indem die Transformation literarischer Genres anhand einer möglichst großen Menge nicht-kanonisierter Werke analysiert wird. Die Fragestellungen, die an diese umfassende Menge von Texten herangetragen werden, wurden aber offensichtlich an dem einen Prozent kanonisierter und genau gelesener Werke entwickelt – ein Zirkel, der die Literaturwissenschaft um potenzielle innovative Fragestellungen bringt. Die literaturwissenschaftliche Gretchenfrage lautet also: Welches sind die dem Gegenstand – den literarischen Texten in ihrer Masse – adäquaten, innovativen Fragestellungen und Suchbegriffe?

Auf andere Weise als von Moretti wird diese Frage aus der Sicht des bereits erwähnten *New Historicism* bzw. der Kulturpoetik beantwortet. Wie dessen Konzept des *Distant Reading* geht auch die Kulturpoetik von einer radikalen

¹¹ Franco Moretti, *Distant Reading*, Konstanz 2016 (engl. 2013); vgl. auch meine Rezension: „Comeback der Philologie? In seiner nun ins Deutsche übersetzten Essaysammlung ‚Distant Reading‘ entwirft Franco Moretti eine literaturgeschichtliche Methodologie für das digitale Zeitalter“, in: *literaturkritik.de*, 9. September 2016, <https://literaturkritik.de/id/22507> (letzter Aufruf 24.9.2017), Thema: Debatte Pro und Contra Distant Reading; zus. m. Dieter Lamping.

¹² Franco Moretti, *Kurven, Karten, Stammbäume. Abstrakte Modelle für die Literaturgeschichte*, Frankfurt a.M. 2009, 10.

¹³ Moretti (Anm. 12), 34.

¹⁴ Moretti (Anm. 11), 64.

¹⁵ Dt.: *Die Schlachtbank der Literatur*, Moretti (Anm. 11), 63–86, hier: 84.

Verbreiterung der historischen Materialbasis aus. Allerdings geschieht dies zum einen nicht so sehr im Hinblick auf Fragen wie die Transformation von Genres, sondern zum Zweck der kulturellen Kontextualisierung von Texten. Zum anderen ist – trotz der Größe des Archivs – nicht der makro-, sondern der mikroskopische Blick entscheidend. „Der New Historicist“, so formuliert Moritz Baßler in seiner 2005 erschienenen Untersuchung *Die kulturpoetische Funktion und das Archiv. Eine literaturwissenschaftliche Text-Kontext-Theorie*,

will Texte geschichtlich, also in ihrem historischen Kontext lesen, weigert sich jedoch, den historisch-kulturellen Kontext eines Werks anders zu fassen denn in Form weiterer, je partikularer Texte. Genau aus diesem Grund gelingt es ihm, seine Lektüre trotz kulturwissenschaftlicher Ausweitung des Gegenstandsbereiches *close* zu halten, so mikroskopisch genau, wie es eben nur der textanalytische Werkzeugkasten hergibt.¹⁶

Die Lektüre hat also paradoxerweise zugleich ‚distant‘, makroskopisch an einer Masse von Texten einer Kultur *und* mikroskopisch-strukturalistisch zu erfolgen. Von entscheidender Bedeutung ist in diesem Zusammenhang nun, die richtigen Fragestellungen und Suchbegriffe zu wählen. Die Kulturpoetik löst dieses Problem, indem sie, das poststrukturalistische Intertextualitätsmodell kulturwissenschaftlich-intermedial erweiternd, nach bestimmten Paradigmen in Texten sucht, diese aber zunächst in ihrem engeren syntagmatischen Kontext analysiert. Den Ausgangspunkt der Untersuchung bilden irritierende syntagmatische Kombinationen von Paradigmen, überraschende Kontexte in einem bestimmten Text. Von diesem Knotenpunkt innerhalb *eines* Textes aus werden ähnliche oder auf spezifische Weise abweichende Kookkurrenzen innerhalb anderer Texte des kulturellen Archivs gesucht.

Beide vorgestellten Ansätze, *Distant Reading* wie Kulturpoetik, bedürfen dabei, da sie sich umfangreichen Textkorpora zuwenden, elektronischer Hilfsmittel. Es handelt sich damit um Anwendungsgebiete der *Digital Philology*; durch die gewaltige Datenmenge ändern sich die Fragestellungen, die Möglichkeiten und Methoden der Untersuchung:

Durchsucht man mit technischen Mitteln einen Text von tausend Seiten nach dem Kontext, innerhalb dessen ein Wort verwendet wird, erledigt man dies etwas schneller, als es menschlichen Forschern, die dies mit dem die Zeile entlanglaufenden Bleistift ohne technische Unterstützung tun, möglich wäre. Tut man dasselbe informationstechnisch unterstützt in einer der großen digitalen Textsammlungen, die tausende Bände enthalten, erhält man Ergebnisse, die ohne diese Unterstützung nicht erzielt werden könnten. Dadurch werden Fragestellungen möglich, die ohne das Werkzeug nicht aufkommen können – die aber in weiterer Folge auch neue methodische Überlegungen erfordern können.¹⁷

¹⁶Moritz Baßler, *Die kulturpoetische Funktion und das Archiv. Eine literaturwissenschaftliche Text-Kontext-Theorie*, Tübingen 2005, 20.

¹⁷Manfred Thaller, „Digital Humanities als Wissenschaft“, in: Fotis Jannidis/Hubertus Kohle/Malte Rehbein (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart/Weimar 2017, 13–18, hier: 14.

Distant Reading und Kulturpoetik sind zweifellos wichtige Bezugspunkte für diese methodologischen Überlegungen.

4 „Wie sich die Kombinationen [...] verändern“: Ein Blick auf Modeblogs

Die angedeuteten methodologischen Perspektiven möchte ich am Beispiel des populärkulturellen Phänomens ‚Blog‘ genauer darstellen. Wie Kundenrezensionen, Mitmach-Plattformen und Fan-Fictions ist auch dieses Genre, ein elektronisches Hybrid aus Tagebuch und Journalismus, weit vom Verdacht entfernt, die Tradition der avantgardistischen Experimentalästhetik fortzuführen – das genaue Gegenteil ist zumeist der Fall. Der Blog ist eine populäre Form; wie die Fan-Fiction ist seine Funktion häufig die Selbstverständigung innerhalb einer Community. Experimentelle Blogs wie *Die Dschungel. Anderswelt*¹⁸ von Alban Nikolai Herbst sind die Ausnahme. Bei ihm handelt es sich um ein radikal subjektives Diarium, das der Autor seit beinahe 20 Jahren tagesaktuell online veröffentlicht. In Form eines elektronischen Arbeitstagebuchs macht er seine literarische Produktion als *work in progress* interaktiv zugänglich; neben konsequenter Verlinkung und multimedialen Elementen verfolgt Herbst, der innerhalb des Blogs verschiedene Identitäten und Pseudonyme annimmt, das ästhetische Programm einer Fiktionalisierung der Wirklichkeit unter den Bedingungen der elektronischen Medienkultur; User*innen, die sich auf die Anschlusskommunikation einlassen, werden Teil der fiktionalen *Dschungel. Anderswelt*. Das Internet wird vom Autor vor diesem Hintergrund zum „phantastischen Raum“ erklärt: „Alles wird hier Literatur und der Autor selbst zur literarischen Figur, die mit anderen, teils realen, teils ebenfalls erfundenen Figuren ein Netzwerk aus Avataren Kommunikatoren bildet.“¹⁹

Der weitaus größte Teil der gegenwärtigen Blogosphäre ist ästhetisch bei Weitem nicht so spektakulär wie *Die Dschungel. Anderswelt* – und doch für den/die Literatur- und Kulturwissenschaftler*in nicht weniger aufschlussreich. Analysiert wurden sie bislang zumeist aus linguistischer korpus- bzw. textsortenanalytischer,²⁰

¹⁸ <http://albannikolaiherbst.twoday.net/> (letzter Aufruf 24.9.2017).

¹⁹ Alban N. Herbst, „Phantastische Räume im Internet (1). Aus dem Entwurf des Marburger Vortrags“, in: *Die Dschungel. Anderswelt* (21. November 2012); <http://albannikolaiherbst.twoday.net/stories/phantastische-raeume-im-internet-1-aus-dem-entwurf-des-marburger-vortr/> (letzter Aufruf 24.9.2017)

²⁰ Vgl. exemplarisch Peter Schildhauer, *Textsorten im Internet zwischen Wandel und Konstanz. Eine diachrone Untersuchung der Textsorte Personal Weblog*, Halle 2014, <http://digital.bibliothek.uni-halle.de/ulbhalhs/urn/urn:nbn:de:gbv:3:4-12640> (letzter Aufruf 24.9.2017); Christoph Moss/Jill Heurich, *Weblogs und Sprache. Untersuchung von linguistischen Charakteristika in Blog-Texten*, Wiesbaden 2015; Joanna Pedzisz, *Profil des Online-Diskurses in Blog-Interaktionen an der Schnittstelle zwischen theoretischem Konzept und empirischem Modell*, Frankfurt a.M. 2017.

diskurstheoretischer²¹ und literatursoziologischer²² Perspektive. Auch für aktuelle Forschungsgebiete wie Populärkultur oder Konsumästhetik stellen Blogs umfassendes Material bereit. Dabei ist eine möglichst große Spannbreite an Formen in den Blick zu nehmen – von Literaturblogs über Mode- und Lifestyleblogs bis hin zu Sport-, Hobby- und Reiseblogs. Gerade eine Beschäftigung mit diesen Formen verspricht, die ‚soziale Energie‘ – die populären Diskurse einer Zeit – in den Blick zu bekommen.

Konstitutiv für Blogs, wie ja bereits für die traditionellen Vorläufer ‚Tagebuch‘ und ‚Tagesjournalismus‘, ist ihre zumeist hohe Frequenz – im Idealfall verfasst ein*e Blogger*in täglich oder mehrmals täglich neue Einträge. Blogs besitzen also, auf welche Weise auch immer, jene von Baudelaire benannten Insignien der Modernität, „das Vergängliche, das Flüchtige, das Zufällige“.²³ Gerade die scheinbare Beliebbarkeit und Kurzlebbarkeit, die der Kommunikation im Internet gerne vorgeworfen wird, gilt es in den Blick zu nehmen. Im Folgenden konzentriere ich mich daher auf eine Form des Blogs, deren inhaltliche Ausrichtung exakt diesem Muster entspricht und es somit gewissermaßen verdoppelt: den Modeblog. Der Anfang dieses Genres wird auf das Jahr 2002 datiert,²⁴ 2015 wurde die Zahl der Modeblogger*innen allein in Deutschland auf 5000 geschätzt, wobei etwa 200 von ihnen von den finanziellen Einnahmen aus dem Bloggen leben konnten.²⁵

An Modeblogs sind zentrale Tendenzen jener fundamentalen medien- und kulturgeschichtlichen Veränderungen abzulesen, die das Internet mit sich bringt. Hin gewiesen wurde darauf bereits 2010 in einem *ZEIT*-Artikel unter dem Titel *Mode als Dialog. Wie das Internet Modekonsumenten zu Gestaltern macht*.²⁶ Bereits die Überschrift zielt auf das für das Medium konstitutive Moment der Interaktion: „In allen möglichen Bereichen hat das Internet Menschen von passiven Konsumenten zu Gestaltern gemacht. Auf Facebook werden sie zu Netzwerknern, eBay macht sie zu Händlern – und die Modeblogs haben in den letzten Jahren unzählige Stilkritiker hervorgebracht.“²⁷ Damit verläuft Mode nicht mehr, diskursiven Exklusionsmechanismen folgend, primär in eine Richtung, von den Modeschöpfer*innen und -firmen, vermittelt über Modejournalist*innen, hin zu den Konsument*innen;

²¹ Vgl. exemplarisch Jenny Lüders, *Ambivalente Selbstpraktiken. Eine Foucault'sche Perspektive auf Bildungsprozesse in Weblogs*, Bielefeld 2007.

²² Vgl. hierzu grundlegend insbesondere Peer Trilcke, „Ideen zu einer Literatursoziologie des Internets. Mit einer Blogotop-Analyse“, in: *Textpraxis* 7/2 (2013), <http://www.uni-muenster.de/Textpraxis/peer-trilcke-literatursoziologie-des-internets> (letzter Aufruf 24.9.2017).

²³ Baudelaire (Anm. 5), 226.

²⁴ Tillmann Prüfer, „Modeblogs. Mode als Dialog“, in: *DIE ZEIT* 8 (2010), <http://www.zeit.de/2010/08/Modeblog> (letzter Aufruf 24.9.2017).

²⁵ Wlada Kolosowa, „Modeblogger. Kann man denn davon leben?“, in: *DIE ZEIT* 3 (2015), <http://www.zeit.de/2015/03/modeblogger-geld-fashion-week> (letzter Aufruf 24.9.2017).

²⁶ Der Artikel von Tillmann Prüfer (Anm. 24) wurde am 18.2.2010 veröffentlicht.

²⁷ Ebd.

vielmehr partizipieren diese selbst aktiv und öffentlich gut sichtbar am Diskurs, indem sie in Blogs Mode thematisieren: „Früher war Mode ein Monolog, jetzt ist sie ein Dialog.“²⁸ Damit ändert sich zugleich – wiederum medientypisch – die Geschwindigkeit. Früher, so der Verfasser des *ZEIT*-Artikels Tillmann Prüfer, hatte „die Modeindustrie [...] ihren eigenen Rhythmus. Sie bestimmte, wann der Stil zum Menschen kam. Sie bestimmte, wer die Entwürfe sehen und damit auch, wer sie bewerten durfte.“²⁹ Mit der Aufgabe der kommunikativen Exklusionsprinzipien geht die Beschleunigung einher: „Nun haben die Luxusmarken ihr öffentliches Bild nicht mehr allein in der Hand. Sie müssen lernen, mit den Massen zu spielen. Denn die Zeiten, als Geschmack eine Frage von Experten und Hochglanzmagazinen war, sind vorbei. Noch bevor die erste Stilkritik gedruckt ist, wird die Mode in Blogs diskutiert, vergöttert oder verrissen.“³⁰

Die ‚Demokratisierung‘ der Mode geht jedoch noch einen Schritt weiter: Die Blogger*innen berichten nicht nur über neuste Modetrends, sondern schaffen eigene Mode. Indem sie über ihre individuellen Vorlieben, ihren eigenen Stil berichten – und zugleich auf den dazugehörigen Fotografien als ihre eigenen Models fungieren –, wird die Mode zumindest in der Tendenz so vielfältig wie die Masse der Akteur*innen, die über sie berichten. Abwechslungsreich sind dabei nicht nur die individuellen Stile, vielfältig ist auch jeder Stil in sich:

Die Blogger spiegeln nicht nur die Mode, sie schaffen ein neues Bild von ihr. [...] [Sie] bilden nicht einfach Trends ab. Sie zeigen Leute, die ihre eigene Mode schaffen: Kompositionen aus Designerteilen, Secondhand- und Kaufhausware. [...] Die Jacke von *Comme des Garçons* wird zur H&M-Jeans getragen, das Gucci-Jackett zu Schuhen von Converse. Die Botschaft: Jeder ist sein eigener Trend.³¹

Man könnte, in einer etwas vorschnellen Analogie, auch sagen: Jedes Model wird zu einer Ansammlung von Links, zu einem Knotenpunkt von Diskursen. Ist Mode – auf den drei Ebenen der realen Kleidung, der textuellen Beschreibung und der fotografischen Wiedergabe bzw. Inszenierung, wie bereits Roland Barthes gezeigt hat,³² ein überaus komplex codiertes syntaktisch-semantisches Phänomen, so entspricht die an den Individualtrend-Blogs abzulesende Komplexitätssteigerung derjenigen des Mediums ‚Internet‘: Alles ist mit allem verknüpfbar. Wie die Kleidung, so folgen dabei auch die einzelnen Blogeinträge den Prinzipien ‚Crossover‘ und ‚Patchwork‘. Das möchte ich an einigen Beispielen erläutern.

Werfen wir zunächst einen Blick auf den Blog *Journelles*. Er ist nach eigenen Angaben

²⁸ Ebd.

²⁹ Ebd.

³⁰ Ebd.

³¹ Ebd.

³² Roland Barthes, *Die Sprache der Mode*, Frankfurt a.M. 10/2017; vgl. ferner Dagmar Venohr, *medium macht mode. Zur Ikonotextualität der Modezeitschrift*, Bielefeld 2010.

das grösste deutschsprachige Modeblogazine und liefert täglich aktualisiert die schönsten News und Inspirationen aus der Modewelt, nimmt die Leser mit auf Reisen um den ganzen Globus und beschäftigt sich mit den wichtigsten Beauty- und Wohn-Trends. Mit dem Autoren-Netzwerk bestehend aus Journalisten, Einkäufern und Designern liefert *Journelles* Outfits, Karriere-Interviews, Interior-Inspirationen und aktuelle Trendjournale.³³

Der Perspektivenreichtum (im Sinne einer hohen Anzahl an Beiträgerinnen),³⁴ aber auch der intendierte Charakter der ‚Normalität‘ und Alltäglichkeit werden besonders in der Rubrik *JOURgarderobe: Closet Diary* deutlich. Hier wird die klassische Form des Tagebuchs aufgegriffen, eine Woche lang berichtet eine Gastbeiträgerin Tag für Tag über ihr Privat- und Arbeitsleben sowie insbesondere über die getragene Kleidung. Das Ganze ist durch Modeaufnahmen reich bebildert, und unter jedem Eintrag findet sich nochmals die Liste der getragenen Marken: „Tasche: Fendi (via Vestiaire Collective), Schuhe: Céline (via Vestiaire Collective), Hose: Maison Margiela, Shirt: Vintage, Jacke: Levi’s (Vintage)“.³⁵

Dass es mit dem Perspektivenreichtum (im Sinne eines tatsächlich andersartigen Blickwinkels) allerdings nicht ganz so weit her ist, geht daraus hervor, dass auch die Gastbeiträgerinnen eng mit der Branche verknüpft sind. In diesem Fall stammt die Produktliste von der „40-jährige[n] Siliva Philipp. [Sie] [...] ist Inhaberin und Geschäftsführerin der Agentur Philipp Seine Helden in Köln. Mit ihrem 20-köpfigen Team betreut sie namenhafte [sic] Kunden aus dem Beauty- und Lifestylebereich wie z. B. L’Oréal oder Schwarzkopf und deckt vom Foto-shooting bis hin zum Packaging-Design über Social-Media sämtliche Bereiche der Kommunikation ab.“³⁶ Entsprechend verläuft ihr Tag zwischen Team-besprechungen, Kundenterminen, Fotoshooting, Fitness-Studio, Netflix, Sushi und Baby Calamari. Was aber verbirgt sich hinter der Produktliste für den Dienstag-Eintrag ihres Wochen-Tagebuchs? Hier der Volltext:

Mein Look an diesem sehr normalen Dienstag? Alltagsauglich, aber mit Akzenten. Das Sesamstraßen-T-Shirt habe ich während eines Türkei-besuches bei Freunden gekauft. Damals habe ich noch studiert. Bis heute trage ich es unglaublich gerne und beobachte über die Jahre, wie sich die Kombinationen damit verändern. Die Jeansjacke habe ich in einem Secondhand-Laden in Berlin gekauft.³⁷

³³ <https://www.journelles.de/about/> (letzter Aufruf 24.9.2017).

³⁴ Die hohe Anzahl der explizit aufeinander Bezug nehmenden Beiträgerinnen sowie die (Anschluss-)Kommunikation im Rahmen des Blogs verstärken das bereits hinsichtlich traditioneller Modezeitschriften zu konstatierende „Freundinnen-Image“, den Charakter eines Lebenshilfe bietenden „vertrauten Kommunikationspartner[s]“ (Hanspeter Ortner, *Wortschatz der Mode. Das Vokabular der Modebeiträge in deutschen Modezeitschriften*, Düsseldorf 1981, 23, <https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/1771> [letzter Aufruf 24.9.2017]).

³⁵ Alexa von Heyden, „Jourgarderobe: Closet Diary mit Silvia Philipp, Inhaberin einer Kölner Werbeagentur“, in: *Journelles* (1. Mai 2017), <https://www.journelles.de/jourgarderobe-close-silvia-philipp/> (letzter Aufruf 24.9.2017).

³⁶ Ebd.

³⁷ Ebd.

Alles ist dezidiert normal und alltäglich, aber „mit Akzenten“, folgt also dem Grundprinzip ‚Norm plus Abweichungen‘: Zu den aktuellen Marken *Fendi*, *Céline*, *Maison Margiela* und *Vintage* kommt als Störfaktor die *Sesamstraße* hinzu (s. Abb. 1).

Sie öffnet auch zeitlich eine zusätzliche Ebene, indem das Sesamstraßen-T-Shirt auf die eigene Studienzeit zurückverweist (wie die Secondhand-Jeansjacke auf eine frühere Besitzerin), also im Gegensatz zu den Markenartikeln nicht in der platten Gegenwart, im konsumästhetischen Netz aktueller Marken verhaftet bleibt, sondern autobiographische ‚Tiefe‘ verströmt. Die *Sesamstraße* ist das Bleibende, der – hinter die Studienzeit auf die Kindheit zurückweisende – kontinuierkeitsstiftende Bezugspunkt, während alles andere sich in stets neuen „Kombinationen [...] veränder[t]“. Der Individualstil mischt hier also nicht nur Verschiedenartiges, er blendet auch Neues, Aktuelles über die Tiefe der Zeit, über Populärkultur-Klassiker. Dabei liefert gerade nicht das aufregend Aktuelle, sondern das Altbewährte die „Akzente“. Als aus der relativen Homogenität aktueller Marken herausfallender Störfaktor ist die *Sesamstraße* im Kontext des *Closet Diary* ein Beispiel für jene *überraschenden syntagmatischen Kombinationen von Paradigmen*, die für die kulturpoetische Untersuchung relevant sind. Die spezifische

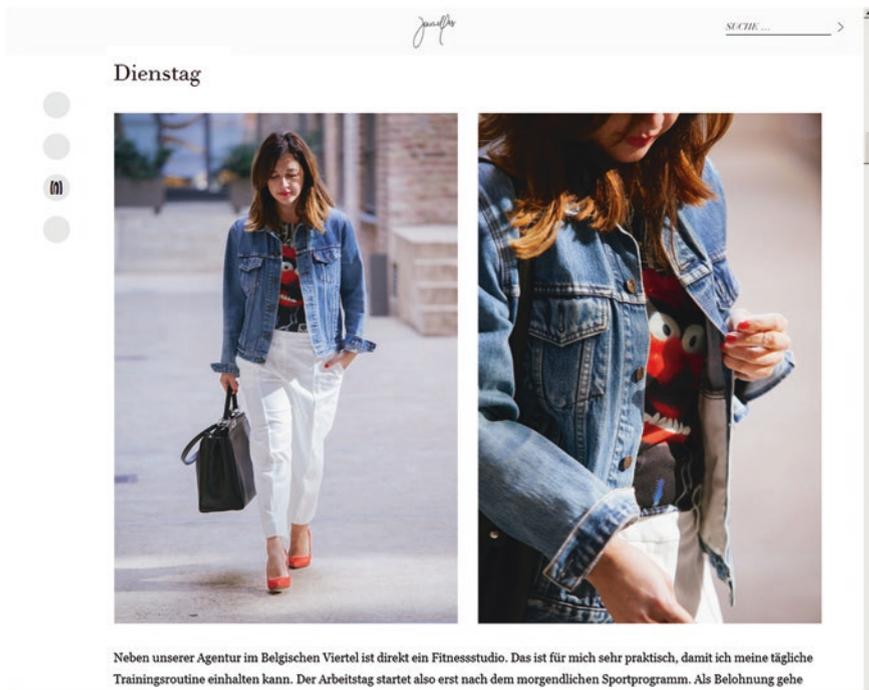


Abb. 1 „Alltagstauglich, aber mit Akzenten“: Das *Sesamstraßen*-T-Shirt (<https://www.journelles.de/jourgarderobe-close-silvia-philipp/>)

Verknüpfung von *Fendi*, *Céline*, *Maison Margiela* und *Vintage* auf der einen sowie der *Sesamstraße* auf der anderen Seite bildet den Knotenpunkt innerhalb dieses Texts, von dem aus nach ähnlichen – und deshalb nicht mehr ganz so überraschenden – Irritationsmomenten in anderen Texten gesucht werden kann.

In der Tat erzielt der Suchbegriff ‚Sesamstraße‘ nicht nur in diesem Modeblog zahlreiche Treffer. Zumeist tritt sie dabei nicht als T-Shirt-Motiv auf, vielmehr rufen bestimmte Kleidungsstücke entsprechende Assoziationen hervor: „Meine neue Lieblingsjacke von Stine Goya lässt mich wie frisch aus der Sesamstraße gehüpft aussehen, aber ich liebe das Teil! Und es ist soooo flauschig. Dazu trage ich Nike Air Force 1, Topshop Jeans, Hut von Maison Michel und eine Tasche von Chloé.“³⁸ Was hier nur erneut bewusster Irritationsfaktor innerhalb des Patchworks ist, besitzt in einem weiteren Eintrag wiederum eine zeitliche Tiefendimension:

Manche Kleider erwecken sofort Kindheitserinnerungen. So geschehen bei meinem Kleid von Carven. Wer muss nicht an Grobi aus der Sesamstraße denken? Einer meiner Lieblingscharaktere und deshalb natürlich auch sofort mein Lieblingskleid.

Heute ist es noch zu heiß, um es zu tragen. Es wird aber in den nächsten Wochen zum Einsatz kommen. Wie Jessie freue ich mich auch auf ein paar kühlere Tage, denn das Leben mit Babybauch ist bei den ungewöhnlich heißen Temperaturen doch ein wenig beschwerlicher. In jedem Falle wird das Sesamstraßenkleid samt Kugel in nächster Zeit am liebsten mit meinen Sneaker [sic] von Ports1961 ausgeführt. Wenn das Baby von innen gucken könnte, würde es sich sicher freuen und leise summen: Mana Mana.³⁹

Die Kontinuität erstreckt sich hier von der eigenen Kindheit über das aktuelle Kleid von Carven (Abb. 2) bis hin zur sich im embryonalen Stadium befindenden folgenden (Kinder-) Generation. Der/die Leser*in des Blogs seinerseits/ihrerseits wird so sehr in den Diskurs einbezogen, dass er/sie am Ende des Beitrags den Mana-Mana-Song als YouTube-Video abspielen kann.

In einem weiteren Eintrag identifiziert sich die Bloggerin mit dem „Reporter-schaft Wolle aus der Sesamstraße! Auffällig oft lassen sich meine Outfits in jüngster Zeit mit Figuren aus der Kindersendung vergleichen.“⁴⁰ Neben Kindheit und nostalgisch-anachronistischem Populärkult-Status darf hinsichtlich der *Sesamstraße* auch der Diskurs der Homosexualität nicht fehlen. So wird im Modeblog *Jane Wayne* – „ein Blogzine für Modeverliebte, für Träumer*innen, junge Frauen, die vor allem sich selbst gefallen wollen, statt dem Rest der Welt, für all jene, die gern entdecken, diskutieren, hinterfragen und staunen“⁴¹ – im August 2011

³⁸ Jessie Weiß, „Jourgarderobe: Closet Diary“, in: *Journelles* (17. November 2014), <https://www.journelles.de/jourgarderobe-closet-diary/> (letzter Aufruf 24.9.2017).

³⁹ Kerstin Görling, „Jourlook: Sesamstraßenkleid von Carven“, in: *Journelles* (16. September 2016), <https://www.journelles.de/jourlook-sesamstrassenkleid-von-carven/> (letzter Aufruf 24.9.2017).

⁴⁰ Jessie Weiß, „Jourlook: Flausch!“, in: *Journelles* (20. August 2013), <https://www.journelles.de/jourlook-flausch/> (letzter Aufruf 24.9.2017).

⁴¹ <http://www.thisisjanewayne.com/> (letzter Aufruf 24.9.2017).

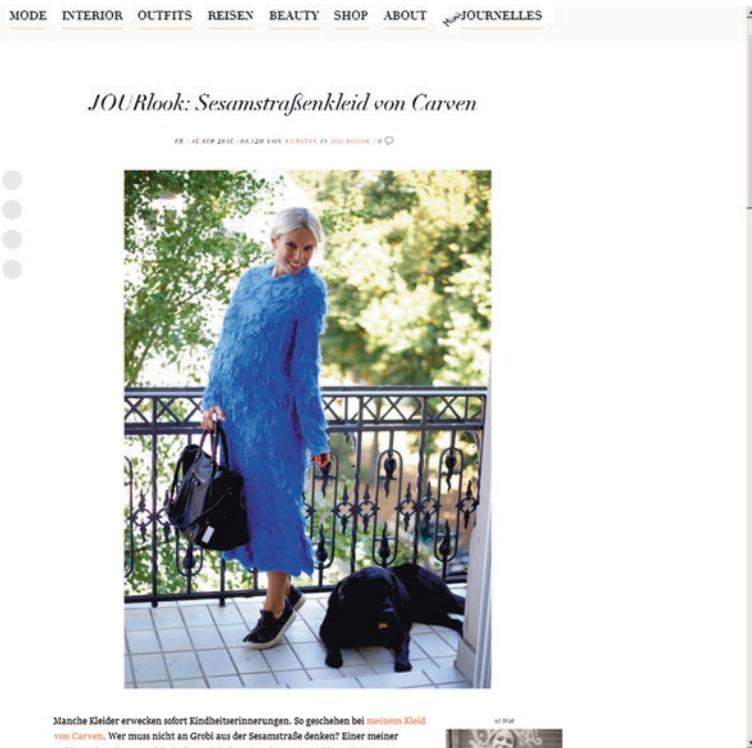


Abb. 2 Mana Mana, generationenübergreifend. (<https://www.journalles.de/jourlook-sesamstras-senkleid-von-carven/>)

über die Petition des „Schwulen-Aktivist[en] Lair Scott“ berichtet, der „von den Sesamstraßen-Machern forderte, Ernie und Bert mögen sich nun endlich outen.“⁴²

Auf eine Metaebene gehoben, wird der *Sesamstraßen*-Diskurs im Modeblog *C'est Clairette*⁴³ der auch für die *FAZ* und *DIE ZEIT* publizierenden Modejournalistin Claire Beermann. Auch sie posiert mit einem „Oberteil [...], in dem ich aussehe wie eine Figur aus der Sesamstraße.“⁴⁴ (s. Abb. 3).

Wieder geht es im Fall dieses Individualstils um eine bewusste Ästhetik der Abweichung: Das Oberteil

⁴² „Keine Heirat für Ernie und Bert – Das Aus der Sesamstraßen-Petition?“, in: *This is Jane Wayne* (15. August 2011), <http://www.thisisjanewayne.com/news/2011/08/15/keine-heirat-fur-ernie-und-bert-das-aus-der-sesamstrasen-petition/> (letzter Aufruf 24.9.2017).

⁴³ <http://cestclairette.com/> (letzter Aufruf 24.9.2017).

⁴⁴ Claire Beermann, „Schlaghosen sind wie Schlagsahne. Die brandneue Klassiker-Kolumne auf C'est Clairette“, in: *c'est clairette* (20. Oktober 2014), <http://cestclairette.com/2014/10/schlaghosen-sind-wie-schlagsahne/> (letzter Aufruf 24.9.2017).

Q |   

c'est clairette *Mode Zeit Leben Spaß Kultur*

ÜBER C'EST CLAIRETTE 

20. Oktober 2014 • **Mode**

 FACEBOOK
 TWITTER
 EMAIL

Neulich habe ich mir ein Oberteil gekauft, in dem ich aussehe wie eine Figur aus der Sesamstraße. Es ist blau, aus Jeansstoff, gefranst, asymmetrisch geschnitten, hat ein Vermögen gekostet und setzt alle Regeln des guten Geschmacks außer Gefecht. Ich musste es haben, es war so schön anders. Mal was Neues. Nach diesem Motto versuche ich immer wieder meine irrationalen Einkäufe zu rechtfertigen. Viel Geld für etwas total Ausgefallenes auszugeben ist meiner These nach nichts weiter als schlaues Investment in eine brillante Innovation, kurz bevor sich die Massen darauf stürzen.



Dieses Dilemma beschäftigt mich immer wieder: welche Investition hat mehr Wert? Die Innovation oder der Klassiker? Das Verrückte oder das Traditionelle? Klassisch – was heißt das überhaupt? Im New Yorker MoMa hängen im 5. Stock Gemälde von Willem de Kooning, der abstrakte Pinselstriche mit gefletschten Zähnen gemalt und das Ganze dann mit so scheinheiligen Titeln wie Woman versehen hat. Diese Bilder bezeichnen wir heute als Klassiker – nicht klassisch im Sinne einer antiken Zeus-Statue, aber klassisch, weil sie verschiedene Stilphasen überdauert und mit den Jahrzehnten immer mehr an Bedeutung gewonnen haben. So wie die Pelzjacke

Auch noch spannend

-  **Mode ohne Drama**
13. September 2017
-  **Schön öko**
4. August 2017
-  **Vom Laufsteg gelernt, Vol. II**
16. März 2017
-  **Verboten ist jetzt erlaubt**
25. Januar 2017
-  **Das verlorene Kleid**
21. Dezember 2016

Abb. 3 „Wie eine Figur aus der Sesamstraße“. (<http://cestclairette.com/2014/10/schlaghosen-sind-wie-schlagsahne/>)

ist blau, aus Jeansstoff, gefranst, asymmetrisch geschnitten, hat ein Vermögen gekostet und setzt alle Regeln des guten Geschmacks außer Gefecht. Ich musste es haben, es war so schön anders. Mal was Neues. Nach diesem Motto versuche ich immer wieder meine irrationalen Einkäufe zu rechtfertigen. Viel Geld für etwas total Ausgefallenes auszugeben ist meiner These nach nichts weiter als schlaues Investment in eine brillante Innovation, kurz bevor sich die Massen darauf stürzen. Dieses Dilemma beschäftigt mich immer wieder: welche [sic] Investition hat mehr Wert? Die Innovation oder der Klassiker? Das Verrückte oder das Traditionelle? Klassisch – was heißt das überhaupt? Im New Yorker MoMa hängen im 5. Stock Gemälde von Willem de Kooning, der abstrakte Pinselstriche mit gefletschten Zähnen gemalt und das Ganze dann mit so scheinheiligen Titeln wie Woman versehen hat. Diese Bilder bezeichnen wir heute als Klassiker – nicht klassisch im Sinne einer antiken Zeus-Statue, aber klassisch, weil sie verschiedene Stilphasen überdauert und mit den Jahrzehnten immer mehr an Bedeutung gewonnen haben. So wie die Pelzjacke meiner Großmutter, die sie sich in den 70er Jahren maßschneidern ließ und die ich neulich mal überwerfen durfte. Wird meine Enkelin irgendwann wohl auch mein gefranstes Jeans-Oberteil ausprobieren?⁴⁵

⁴⁵ Ebd.

Das Kleidungsstück setzt die „Regeln des guten Geschmacks“ außer Kraft, es ist „so schön anders“, indem es auf einen anderen Diskurs anspielt, anstatt dem Code aktueller Modemarken dem Retro-Code einer Jahrzehnte alten Kinder-Fernsehserie gehorcht.

Für Claire Beermann resultieren daraus grundsätzliche modetheoretische Fragestellungen:

Wieso werden manche Dinge – Gemälde von de Kooning, Pelzjacken, schwarze Blazer, Nutella, Zahnpasta von Elmex, Rennwagen von Porsche, Plattenspieler, Yves-Saint-Laurent-Trenchcoats, Bonnie & Clyde – zu Klassikern und andere nicht? [...] Und wo liegt eigentlich der Unterschied zwischen klassisch und *normcore*? Sind Nutella, Elmex Zahnpasta, *steak frites* und schwarzer Blazer neben ihrer Tätigkeit als Klassiker nicht auch irgendwie total normal?⁴⁶

Das Spannungsfeld besteht hier aus den Begriffen ‚innovativ‘, ‚klassisch‘ und ‚normal‘. Das Innovative hat möglicherweise das Zeug zum Klassiker, ohne dass ein solcher Erfolg voraussagbar wäre, während das Klassische Gefahr läuft, schlicht konventionell zu sein. Claire Beermann reflektiert die Frage, die im Fall des *Sesamstraßen*-T-Shirts der Werbeagentin Siliva Philipp implizit blieb; hier werden die drei Kategorien übereinander geblendet: Die *Sesamstraße* ist ein normaler Klassiker, während ein Oberteil, das aussieht wie aus der *Sesamstraße*, erstens ironisch an diesem Status partizipiert, zweitens dadurch irritiert, dass es dieses Zitat in den ‚falschen‘ Kontext aktueller Mode setzt, und drittens genau durch diese Irritation so innovativ sein möchte, dass es zum Klassiker werden könnte.

Die *Sesamstraße* wirkt im Kontext von Modeblogs als Irritationsmoment; wir stoßen auf die merkwürdige syntagmatische Verknüpfung von Paradigmen, die höchst unterschiedlichen Diskursen entstammen. Nun nennt Claire Beermann in ihrem von der *Sesamstraße* ausgehend das Klassiker-Problem diskutierenden Beitrag *Schlaghosen sind wie Schlagsahne* aber eine ganze Reihe weiterer Paradigmen, die, zwischen Normalität und Klassiker-Status changierend und dem Innovationsanspruch der Mode widersprechend, auf einer Linie mit der Kinderreihe liegen, darunter die Marken⁴⁷ *Elmex* und *Nutella*.

Interessieren wir uns für irritierende Verknüpfungen, so haben wir durch diese Ansammlung von Schlagwörtern in einem Modeblog Suchbegriffe für die weitere Beschäftigung mit dieser Textsorte erhalten. Wir gelangen von der Kombination ‚Mode, *Sesamstraße* und *Nutella*‘ im einen Text etwa über den Suchbegriff

⁴⁶ Beermann (Anm. 44).

⁴⁷ Zur Semiotik und Poetik der Marke sowie zur Warenästhetik vgl. aus literatur- und kulturwissenschaftlicher Perspektive Thomas Wegmann, *Dichtung und Warenzeichen. Reklame im literarischen Feld 1850–2000*, Göttingen 2011; Heinz Drügh/Christian Metz/Björn Weyand (Hg.), *Warenästhetik. Neue Perspektiven auf Konsum, Kultur und Kunst*, Berlin 2011; Björn Weyand, *Poetik der Marke. Konsumkultur und literarische Verfahren*, Berlin/Boston 2013.

„Nutella“ zu weiteren merkwürdigen Befunden der Kookkurrenz von Marken-Nostalgie und modischem Trend. Auch hier kann die Verknüpfung sehr unauffällig daherkommen, so im Fall der *Closet Diary*-Gastbeiträgerin Kathrin Bommann, einer „Produktmanagerin und Einkäuferin“, die zu berichten weiß, sie frühstücke „im Büro erstmal ein feines Brötchen mit Nutella, natürlich!“⁴⁸

Auch dieser Marken-Klassiker wirkt als diskursiver Knotenpunkt. Zum einen assoziieren die Bloggerinnen damit eine heile Welt der Kindheit, in der alles noch weniger kompliziert und überschaubarer war. Anlass dazu gibt im Blog *Jane Wayne* die im Rahmen eines Produkt-Adventskalenders vorgenommene Präsentation des *Samsung Serif TV*, eines Fernsehers, „der durch seinen geschwungenen Körper schon jetzt zum Klassiker avanciert: Ohne qualitative Einbußen, dafür mit ganz viel Gespür, mit allem, was solch ein Apparat heute können muss und mit bestem Verständnis für Design im Gepäck.“⁴⁹ Das Gerät weckt zugleich Erinnerungen an die Kindheit:

Früher einmal, da gehörte ein Fernseher zur Grundausstattung eines jeden Wohnzimmers, war mehr als bloß Flimmerkiste, sondern vielmehr Statussymbol, Berieselungsmaschine und Unterhaltungsmittelpunkt zum Anknipsen. Er war Familienmagnet, manchmal sogar -Mitglied [sic], und sorgte verlässlich jeden Samstagabend dafür, dass wir es uns bepackt mit unseren Nutella-Brotten auf dem Sofa gemütlich machten: Der Fernseher, ach der Fernseher. Irgendwann allerdings wurden wir groß, die treue Röhre immer flacher und das TV-Programm hielt auch nicht mehr das, was uns die Hörzu einst versprach.⁵⁰

Nutella wird hier zum Symbol der Regression, der Sehnsucht nach einem verlorenen medialen Mittelpunkt, nach einer Zeit, in der die Familie magnetisch zusammengehalten wurde und die *Hörzu* verlässlich Rat gab – deutlicher könnte der Gegensatz zur absoluten Vielstimmigkeit des Internet nicht sein.

Weitere Diskurse, mit denen die Marke verbunden wird, sind Gesundheit und Figur. So heißt es auf *Jane Wayne* unter der Rubrik *Beauty* in einem Beitrag mit dem Titel *Mein Masterplan für reine Erwachsenenhaut*, den Zusammenhang von ungesunder Ernährung und Hautproblemen könnten „wir auch gut an uns selbst beobachten, wenn wir mal wieder zu tief ins Nutellaglas geschaut haben“.⁵¹ Auch im Blog *Journelle* wird mehrfach die Beziehung zwischen *Nutella* und weiblichen

⁴⁸ [o.N.], „OURgarderobe: Closet Diary mit Kathrin Bommann, Produktmanagerin und Einkäuferin“, in: *Journelles* (20. März 2017), <https://www.journelles.de/jourgarderobe-closet-diary-kathrin-bommann/> (letzter Aufruf 24.9.2017).

⁴⁹ Sarah Gottschalk, „Adventskalender. Tor 1. Der Samsung Serif TV by Ronan & Erwan Bouroullec im Wert von 1000 Euro“, in: *This is Jane Wayne* (1. Dezember 2016), <http://www.thisisjanewayne.com/news/2016/12/01/adventskalender-2016-tor-1-der-samsung-serif-tv-by-ronan-erwan-bouroullec-im-wert-von-1000-euro/> (letzter Aufruf 24.9.2017).

⁵⁰ Gottschalk (Anm. 48).

⁵¹ Scalamari Jane [d.i. Sarah Radowitz], „Beauty. Mein Masterplan für reine Erwachsenenhaut“, in: *This is Jane Wayne* (10. Juni 2016), <http://www.thisisjanewayne.com/news/2016/06/10/beauty-mein-masterplan-fuer-reine-erwachsenenhaut/> (letzter Aufruf 24.9.2017).

Figurproblemen und damit wiederum zum Mode-Diskurs hergestellt; unter dem Schlagwort „Clean Eating“ heißt es: „34 gibt’s nicht mit Nutella.“⁵²

Komplexer sind die Verknüpfungen, und damit kommen wir zum Ausgangspunkt unserer Suche nach *Nutella* in Modeblogs zurück, wiederum im Blog *C’est Clairette*. Der Zusammenhang von Mode und ungesundem Essen wird hier ironischerweise gerade umgekehrt, denn „Mode macht hungrig“ – und arm: „Wer sich bei Opening Ceremony in den Ruin gekauft“ habe, müsse eben Fastfood wie mit „Nutella gefüllt[e]“ Teigfladen essen.⁵³ Der Ernährungs- und Schlankheitswahn wird ironisiert, wenn darauf hingewiesen wird, „dass das Leben mit Nutella-Brot und Spaghetti Napoli mehr Spaß macht als ohne“.⁵⁴ Die Kindheit wird als eine Lebensphase dargestellt, in der weibliche Körperideale noch keine Rolle spielten: „Als ich 5 Jahre alt war, naschte ich mittags nach dem Kindergarten aus dem Nutella-Glas und trank Vollmilch mit Himbeersirup. Wie meine Figur aussah, interessierte mich überhaupt nicht.“⁵⁵ Es geht somit weniger um die regressive Sehnsucht nach einer heilen Kindheit als vielmehr um die Ablehnung von Geschlechter-Stereotypen und Körperidealen. Die Bedeutung als Rettungsanker erhält die Marke hier interessanterweise in der Phase der Adoleszenz, im spannungsvollen Zusammenspiel mit Koffein und Schmerzmedizin: „I began drinking coffee regularly around the time I started going out and having hangovers, which I cured with the magic combination of one Ibuprofen, two Nutella toasts and three cups of coffee so strong a spoon could stand erect in the mug (almost).“⁵⁶

Die diskursiven Verknüpfungen, die sich ergeben, wenn in Modeblogs eben nicht nur aktuelle Modelabels, sondern auch Marken-Klassiker angeführt werden, möchte ich abschließend noch anhand von zwei Produkten demonstrieren, deren Konnotationen so komplex wie idiotisch sind, indem sie als ‚Südworte‘ einen traditionellen Sehnsuchtsort evozieren: *Capri-Eis* und *Capri-Sonne*. *Capri* ist als Markenbezeichnung, zumal im Zuge des Retro-Kults, wie *Nutella* ein echter Klassiker der deutschen Konsumkultur. Die *Süddeutsche Zeitung* und die Wochenzeitung *DIE ZEIT* erwiesen dem Stieleis in den letzten Jahren in humor-

⁵² Jessie Weiß, „Gesprächsstoff: Auch in der Mode Trend – Detox, Clean Eating und Co“, in: *Journelles* (7. Februar 2014), <https://www.journelles.de/gesprachsstoff-auch-in-der-mode-trend-detox-clean-eating-und-co/> (letzter Aufruf 24.9.2017).

⁵³ Claire Beermann, „Mode macht hungrig. Ein Monat, vier Städte, zwanzig Restaurants“, in: *c’est clairette* (11. Februar 2017), <http://cestclairette.com/2017/02/mode-macht-hungrig/> (letzter Aufruf 24.9.2017).

⁵⁴ Claire Beermann, „Weisheiten 2014. Rückblick auf das Jahr des Grünkohls“, in: *c’est clairette* (30. Dezember 2014), <http://cestclairette.com/2014/12/weisheiten-2014/> (letzter Aufruf 24.9.2017).

⁵⁵ Claire Beermann, „Das neue Blond. Barbie gibt’s jetzt auch in klein und rund. Leider ist sie immer noch zu schön“, in: *c’est clairette* (6. Februar 2016), <http://cestclairette.com/2016/02/das-neue-blond/> (letzter Aufruf 24.9.2017).

⁵⁶ Claire Beermann, „A Farewell Letter to Coffee. This could have been such a sweet love story“, in: *c’est clairette* (18. Juni 2017), <http://cestclairette.com/2017/06/a-farewell-letter-to-coffee/> (letzter Aufruf 24.9.2017).

vollen Artikeln ihre Referenz.⁵⁷ Unter dem Titel *Capri oder Split. Zwei Stieleis-Klassiker sind nach Badeorten in Italien und Kroatien benannt. Wo genießt es sich besser? Ein eiskalter Vergleich* heißt es in der ZEIT:

Whiskytrinker pilgern nach Schottland, Trüffelsucher ins Périgord – wohin aber reist der Freund von Eis am Stiel? Natürlich nach Capri. Oder nach Split. Am besten nach Capri und Split. Das sind seit über fünfzig Jahren die Klassiker auf deutschen Eistafeln und noch immer die großen, die ganz großen Namen in der Welt der Stieleisesser. Wie nennt man solche Leute überhaupt? Stielies? Scheiden sie sich sogar in Untergruppen, in Splitologen und Capriccionados? Ich bin jedenfalls beides und weiß es zu schätzen, dass der Hersteller Langnese Modellpflege betreibt.⁵⁸

Auch diese Markennennungen bewirken neben der Sehnsucht nach Sommer, Sonne und Hitze die Regression in die heile Konsum-Welt der Kindheit, die wiederum auch auf die nächste Generation der eigenen Kinder projiziert wird. Im Blog *Jane Wayne* werden im Rahmen der Kolumne *10 Random Facts about ...* die Zukunftsträume der als Online-Editorin und Social Media Consultant tätigen Scalamari Jane vorgestellt: „Nicht jetzt sofort – aber bald“ wünscht sie sich „ein Haus im Grünen, im bürgerlichen Part von Berlin, ungefähr so, wie und wo ich auch aufgewachsen bin – eine richtig spießige Angelegenheit wird das werden – so mit Hund und mega coolem Kind (auf keinen Fall trilingual erzogen und immer, immer mit Capri-sonne in der Hand)“.⁵⁹

In einem anderen Fall wird *Capri-Sonne* sogar zur literarisch gewagten Metapher für die Jugendliebe:

Atemlos sitze ich am Küchentisch und starre auf das verblasste Foto, das neben all dem dreckigen Geschirr wie eine Oase wirkt. Er hält mich im Arm und ich einen Joint in der Hand, es ist Sommer, zwei Capri-sonnen lehnen aneinander als könnte sie nichts jemals trennen. Noch mal 14 sein.⁶⁰

Für die Modebloggerinnen gehört die *Capri-Sonne* zum Festhalten am Ideal des Mädchenhaften:⁶¹

⁵⁷ o.A., „Capri? Lieber Ischia!“, in: *Süddeutsche Zeitung* (14. Mai 2010), <http://www.sueddeutsche.de/leben/eis-test-capri-lieber-ischia-1.232432> (letzter Aufruf 24.9.2017); Oliver M. Schmitt, „Capri oder Split? Zwei Stieleis-Klassiker sind nach Badeorten in Italien und Kroatien benannt. Wo genießt es sich besser? Ein eiskalter Vergleich“, in: *DIE ZEIT* 34 (2012), <http://www.zeit.de/2012/34/Stieleis-Kroatien-Split-Italien-Capri> (letzter Aufruf 24.9.2017).

⁵⁸ Schmitt (Anm. 56).

⁵⁹ Scalamari Jane [d.i. Sarah Radowitz], „10 Random Facts about ...“, in: *This is Jane Wayne* (11. April 2016), <http://www.thisisjanewayne.com/news/2016/04/11/10-random-facts-about-scalamari-jane/> (letzter Aufruf 24.9.2017).

⁶⁰ Nike van Dintter, „Die Sache mit dem verlorenen Freund“, in: *This is Jane Wayne* (10. März 2013), <http://www.thisisjanewayne.com/news/2013/03/10/die-sache-mit-dem-verlorenen-freund/> (letzter Aufruf 24.9.2017).

⁶¹ Zum Ideal des Kindlichen in der Mode vgl. auch Barthes (Anm. 32), 247.

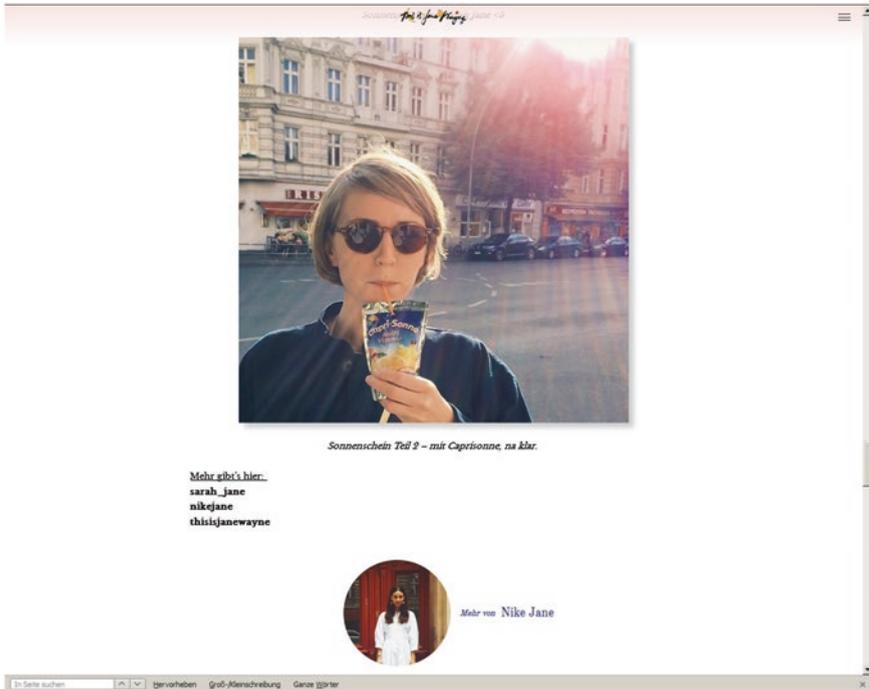


Abb. 4 Die heile Konsum-Welt der Kindheit: Capri-Sonne. (<http://www.thisisjanewayne.com/news/2014/05/13/insta-diary-unsere-bilder-der-woche-4/>)

am Wochenende hatte ich einen Abend nur für mich allein, also tat ich, was ich gerne tue, wenn niemand stört, nämlich im Früher versinken, Fotos rauskramen und alten Freunden schreiben, an wilde Zeiten denken, mich fragen, ob es das Leben mit 17 mehr Sinn macht oder mit 26, Capri-Sonne trinken und Filme schauen, solche, die ich eigentlich längst gesehen haben wollte, aber nie dazu kam, weil das Erwachsenwerden immer wieder dazwischen grätschte. [...] Ich glaube [...], als Mädchen hörst du niemals auf Mädchen zu sein, selbst wenn andere schon „Frau“ zu dir sagen.⁶²

Gerne wird das mädchenhafte Image auch durch Selfies mit *Capri-Sonne* unterstrichen (s. Abb. 4).⁶³

⁶²Nike van Dinther, „Film-Tipp für graue Tage & Soundtrack. ‚17 Filles‘“, in: *This is Jane Wayne* (23. Februar 2015), <http://www.thisisjanewayne.com/news/2015/02/23/film-tipp-fur-graue-tage-soundtrack-17-madchen/> (letzter Aufruf am 24.9.2017).

⁶³Nike van Dinther, „Insta-Diary. Unsere Bilder der Woche“, in: *This is Jane Wayne* (13. Mai 2014), <http://www.thisisjanewayne.com/news/2014/05/13/insta-diary-unsere-bilder-der-woche-4/> (letzter Aufruf 24.9.2017).

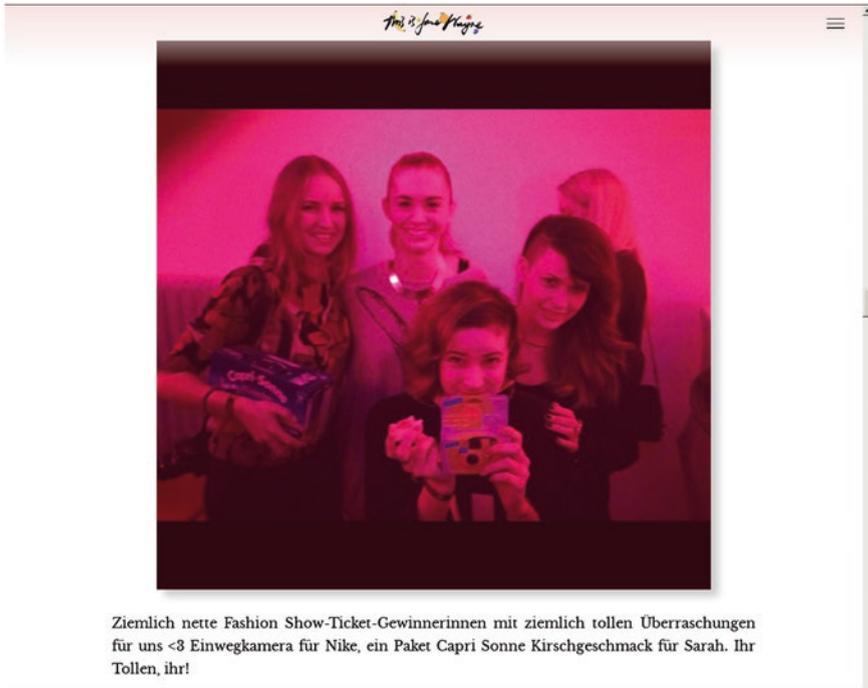


Abb. 5 Glückliche Gewinnerinnen. (<http://www.thisisjanewayne.com/news/2012/11/09/44440/>)

Dass es sich um einen Diskurs mit Überlappungen zur Modeszene handelt, geht daraus hervor, dass eine Bloggerin als „Fashion Show-Ticket-Gewinnerin [...] ein Paket Capri Sonne Kirscheschmack“ erhält (s. Abb. 5).⁶⁴

Im Fall von *Capri*-Eis ist der Bezug zur Mode durch die Konnotation ‚Sommer‘ hergestellt:

Es kitzelt bereits in unserer Nase und kündigt sich lang ersehnt, unaufhaltsam an: Der Sommer steht in den Startlöchern! [...] Denke ich an heiße Sommertage, kommt mir unweigerlich Zitronen-Eis in den Sinn, eisgeköhlte [sic] Limonade, spritziges Capri-Wassereis, kühle Verstecke und luftige Kleidung in den Sinn [sic]. Und noch nie, und damit schließe ich Kindertage an dieser Stelle aus, war mir der Sinn nach Stücke [sic], die Früchte zieren – gängiger waren da schon die zahlreichen Blumenmuster.⁶⁵

⁶⁴ Nike van Dinther, „Erste Schnappschüsse vom Abend: Frontline Shop Fashion Show“, in: *This is Jane Wayne* (9. November 2012), <http://www.thisisjanewayne.com/news/2012/11/09/44440/> (letzter Aufruf 24.9.2017).

⁶⁵ Sarah Gottschalk, „Ganz schön fruchtig“, in: *This is Jane Wayne* (29. März 2011), <http://www.thisisjanewayne.com/news/2011/03/29/ganz-schon-fruchtig/> (letzter Aufruf 24.9.2017).

Manche Modeaccessoires werden direkt mit dem Eis verglichen, so die limitierte Sonnenbrille *Arancia*, „die euch beim Aufsetzen sofort an Capri-Eis, Orangenplantagen, Italien-Urlaub und schönste Sonnenstunden erinnern soll.“⁶⁶

Andere Modeartikel heißen bereits wie das Eis, so die Capri-Hose oder die Brille Modell *Capri* von *Ace & Tate*. Capri steht dabei – in krassem Gegensatz zum Mainstream-Wassereis – für mondänes Leben und Celebrities; so wird im Sommer 2016 mehrfach über die Hochzeit des Immobiliencycoons Oscar Engelbert auf Capri berichtet. Das dient als Vorbild für alle Akteurinnen der Mode-Community, die selbst berühmt werden möchten. Im Rahmen von Tipps, wie man zum Instagram-Star aufsteigt, wird daher großer Wert auf das mondäne Setting gelegt:

Also immer schön überlegen, welche Klamotten vor welcher Kulisse funktionieren könnten. „Der Hintergrund und das Outfit müssen stimmig sein“, schreibt Aimee. So ist auch das folgende Foto vor dem Taj Mahal in einem Rüschenkleid von Anna October und goldenen Jimmy Choo Pumps alles andere als zufällig entstanden, ebenso wenig wie die Streetstyle-Fotos aus Paris, New York oder Capri.⁶⁷

Auch im Blog *C'est Clairette* ist Capri ein Statussymbol⁶⁸ und ein Traumort für das Ideal modischer Leichtigkeit. Spezielle Modeartikel führen zur bewussten Imagination von Anlässen und Orten, an denen sie getragen werden könnten:

Bei nahezu jeder Show, der ich selig beiwohnen darf, fällt mir jeweils eine passende Situation für dieses oder jenes Outfit ein: bei [sic] der glamourös-lässigen Poolparty im Stil der Siebziger [sic] Jahre auf Capri würde ich selbstredend in einer weißen, luftigen Marlenehose aus Seide von Escada Sport, oder vielleicht auch im apricot-farbenen Maxi-kleid von Rena Lange aufkreuzen [...].⁶⁹

⁶⁶Sarah Gottschalk, „It's official: Jane Wayne x Viu Eyewear. Unsere limitierte Sonnenbrille ‚Arancia‘“, in: *This is Jane Wayne* (19. Juli 2017), <http://www.thisisjanewayne.com/news/2017/07/19/this-is-jane-wayne-x-viu-eyewear-unsere-limitierte-sonnenbrille-arancia/> (letzter Aufruf 24.9.2017).

⁶⁷Alexa von Heyden, „Die magische 12, das perfekte Outfit-Foto und die schönsten Klischees. So wird man laut Aimee Song (‚Song of Style‘) zum Instagram-Star“, in: *Journelles* (16. November 2016), <https://www.journelles.de/aimee-song-song-of-style-instagram-star/> (letzter Aufruf 24.9.2017).

⁶⁸Wiederum heterogene Diskurse miteinander verbindend, schreibt sie etwa im Blogeintrag „Die Biomarktfalle. Sind wir jetzt verwöhnt oder nur anspruchsvoll?“. „Komisch ist bloß, dass wir uns mit dem Einkauf im Biomarkt vorgaukeln, etwas Gutes für die Umwelt zu tun, obwohl wir damit am Ende doch nur unser eigenes Streben nach dem nächsthöheren gesellschaftlichen Rang befriedigen, jenem Level, in dem man sich eben nicht mehr über Hausarbeiten, sondern über Ferienhäuser auf Capri unterhält.“ Claire Beermann, „Biomarkt-Falle. Verwöhnt oder anspruchsvoll“, in: *c'est clairette* (11. Juli 2016), <http://cestclairette.com/2016/07/biomarkt-falle-verwoehnt-oder-anspruchsvoll/> (letzter Aufruf 24.9.2017).

⁶⁹Claire Beermann, „Update aus Berlin“, in: *c'est clairette* (5. Juli 2012), <http://cestclairette.com/2012/07/update-aus-berlin/> (letzter Aufruf 24.9.2017).

The screenshot shows the homepage of the blog 'c'est clairette'. At the top, there is a search bar and social media icons for Facebook, Twitter, and Email. The main header features the blog's logo and the tagline 'Mode Stil Leben Spaß Kultur'. Below this, the date '30. Mai 2015' and the title 'Badeanzug auf Abwegen' are displayed. A sub-headline reads 'BERLIN IST EBEN NICHT BRINDISI'. The main image shows a woman in a pink top and blue jeans posing in a swimsuit next to a white statue in an urban setting. Below the image, there is a text block starting with 'Seit ungefähr vier Monaten warte ich auf den Sommer...' and a list of other articles under the heading 'Auch noch spannend'.

Abb. 6 Im Einsatzgebiet. (<http://cestclairette.com/2015/05/badeanzug-auf-abwegen/>)

Diese Assoziationen bleiben erhalten, wenn es wiederum zur Kreuzung zwischen Modeblog und der populären Eismarke *Capri* kommt. Für den/die Kulturhistoriker*in, der/die in einigen Jahrzehnten etwa dem Lebensgefühl des Frühjahrs 2015 zwischen dicht aufeinanderfolgenden Schlechtwetterlagen und Griechenland-Krise auf die Spur kommen möchte, dürften der Blogbeitrag vom 30. Mai 2015 *Badeanzug auf Abwegen. Berlin ist eben nicht Brindisi* und die ihn illustrierenden Fotos der in urbanem Setting demonstrativ im Badeanzug vor einer antiken Statue posierenden Clairette wertvolles Material bieten (Abb. 6).

Dabei ist erneut eine Untersuchung bis auf die Mikroebene einzelner Lexeme notwendig; nicht nur für eine konsumästhetische Untersuchung ist es so von Interesse, dass sich in Claire Beermanns Lifestyle-Blog die paradoxe Formulierung findet, die Zuversicht auf den Sommer sei „geschmolzen wie ein Capri-Eis in der Julisonne“.⁷⁰

Solche augenzwinkernden Spiele mit der Marken- und Konsumkultur verleihen dem Blog den Effekt der Leichtigkeit und eine Aura der Zeitgenossenschaft. Sie laden den Text semantisch auf, indem erneut ein reizvolles Spannungsverhältnis

⁷⁰Claire Beermann, „Badeanzug auf Abwegen“, in: *c'est clairette* (30. Mai 2015), <http://cestclairette.com/2015/05/badeanzug-auf-abwegen/> (letzter Aufruf 24.9.2017).

zwischen Haute Couture und Populärkultur entsteht. Im Fall von *C'est Clairette* kann die glatte Eleganz des Modeblogs nicht darüber hinwegtäuschen, dass der Reiz, den hier die Artikulation von Zeitgenossenschaft ausübt, aus der Heterogenität des verwendeten Wortmaterials resultiert; die Faszination besteht gerade in der glatten Oberfläche, zu der er die unterschiedlichsten Diskursartikel verwebt. Am Schluss des Artikels *Badeanzug auf Abwegen. Berlin ist eben nicht Brindisi* ist so die Rede davon, „das Repertoire an Einsatzgebieten von Badeanzügen radikal zu erweitern“.⁷¹ Das Sprechen über Mode wird also nicht nur mit populärkulturellen Retro-Phantasien verbunden, hinzu kommt ein Vokabular, das auf militärische Auslandseinsätze in Krisengebieten und so auf die zeitgeschichtliche Situation des frühen 21. Jahrhunderts verweist.

Die Verweisstruktur der Begriffe ‚Capri‘ und ‚Capri-Eis‘ wird dabei im Blog selbst reflektiert:

Im Strandbad Wannsee liegt die italienische Riviera – man muss nur dran glauben. Eine erstaunlich großzügige Fläche feinen weißen Pulversandes, getupft mit sorgfältig aufgereihten, blau-weiß gestreiften Strandkörben. Dahinter ein hellbrauner Klinkerbau mit überdachter Promenade und allerlei Geschäften für die kleinen Bedürfnisse des Strandlebens: Gummiboote, Softeis, Pommes rot-weiß. [...] Im Strandbad Wannsee steht die Zeit still; da herrscht so eine merkwürdige Atmosphäre mondäner Schüßigkeit, der braune Stein, die Markisen der Strandkörbe, schmelzendes Capri-Eis in der Sonne, Palmen in Blumentöpfen, ein paar Trauerweiden am Uferrand, Badegäste mit Käseblatt-Lektüre und all das spiegelt im glattglänzenden Gewässer.⁷²

Berlin ist weder Brindisi noch Capri – das im Wannsee-Strandbad schmelzende *Capri-Eis* weist genau auf diese Differenz hin. Dem entspricht exakt das Motiv der eben nur auf den Süden verweisenden, in Blumentöpfen aufgestellten Palmen im Kontrast zu den tatsächlich am Uferrand wachsenden Trauerweiden. Wer sich nach dem Genuss von *Capri-Eis* sehnt, sehnt sich nach dem ausbleibenden Sommer und nach der wunderbaren Konsum-Welt der Kindheit; wer aber *Capri-Eis* konsumiert, *ist* wiederum gerade nicht in Capri. Wir befinden uns in einer endlosen Reihe von Verweisungen, von Aufschüben. Eben darin treffen sich womöglich die Versprechungen der Marken-Nostalgie mit den Versprechungen der Mode – auch wenn es äußerst Gegensätzliches ist, was sich hier trifft.

5 Zeitdiagnostik: Noch einmal zur literaturwissenschaftlichen Methodologie

Rekapitulieren wir noch einmal das Vorgehen in unserer selektiven Blog-Analyse und fragen vor diesem Hintergrund nach dem methodologischen Ertrag, um dessentwillen wir allein nach abstrusen Dingen wie *Sesamstraße*, *Nutella* und

⁷¹ Ebd.

⁷² Claire Beermann, „Sommer vorm Balkon“, in: *c'est clairette* (7. Juli 2014), <http://cestclairette.com/2014/07/sommer-vorm-balkon/> (letzter Aufruf 24.9.2017).

Capri-Eis in Modeblogs gesucht haben. Unsere Frage war, inwiefern es in diesen Texten zur Überkreuzung von Diskursen, etwa des Mode-Diskurses und der Marken-Nostalgie, kommt. Solche Überkreuzungen konnten wir immer wieder feststellen. Gerade sie sagten etwas über die Texte, deren Machart – und über die Zeit, in der diese dezidiert aktuellen Texte situiert sind, aus.

Wenn Georg Simmel zu Beginn des 20. Jahrhunderts prätendierte, den Henkel einer Vase als ästhetik- und kulturgeschichtliches Indiz einer Epoche interpretieren zu können,⁷³ so kann noch der scheinbar belangloseste Diskurspartikel in einem Blog auf seine zeitdiagnostisch-kultursemiotische Relevanz hin untersucht werden. Es ist, ganz im Sinne von Baudelaires ‚Moderne‘-Definition, wohl kein Zufall, dass wir solche Indikatoren für Zeitgenossenschaft gerade in einem Modeblog finden. Wer zeitgenössischen Diskursen und vor allem Diskurs-Verknüpfungen mikroskopisch auf die Spur kommen möchte, sollte also zum Beispiel Modeblogs analysieren.

Was aber bedeuten die Beobachtungen zum spezifischen Umgang mit Sprachmaterial in diesen Blogs für die literaturwissenschaftliche Auseinandersetzung mit Literatur im Internet? Zunächst einmal besteht ein Vorteil ganz pragmatisch in der leichten und massenhaften Verfügbarkeit. Möchte man untersuchen, inwiefern die Verwendung und Verknüpfung von Sprachmaterial etwas über eine Zeit wie die unsere aussagt, so ist dieser pragmatische Vorteil – die Möglichkeit, in einem großen elektronischen Textkorpus zu suchen – von großem Wert. Ohne Suchfunktion wäre es schlicht unmöglich, in der von mir vorgeführten Weise Fragen an umfangreiche Korpora wie über Jahre hinweg geführte Blogs heranzutragen. Insofern handelt es sich im Kontext der *Digital Philology* also um eine Form des *Distant Reading*. Als Konsequenz der skizzierten Entwicklung weg vom elitären literarischen Experiment ermöglicht dies, sich auf adäquate Weise mit der Populärkultur, der ubiquitären literarischen Kommunikation im Netz zu befassen. Gerade in dieser Hinsicht sind literarische Praktiken im Internet in methodologischer Hinsicht interessant: als kulturelle Phänomene, die massenhaft Daten über Produktions- und Rezeptionsprozesse von Texten verfügbar machen.

Untersuchen wir aber das Funktionieren kommunikativer Prozesse, so analysieren wir immer auch die Machart von Texten. Bezeichnenderweise hat sich in unserer exemplarischen Beschäftigung mit Blogs auf diese Weise ein ergiebiges Untersuchungsfeld aufgetan: Symptomatisch erschienen dort nicht nur die Akte der Kommunikation, sondern die Texte in ihrer spezifischen Gemachtheit, die es in kulturdiagnostischer Hinsicht zu analysieren gilt. Ob und in welcher Weise Texte etwas über ihre Zeit aussagen, lässt sich nicht allein durch quantitative Analysen feststellen, so nützlich diese auch sind; von zeitdiagnostischem Wert sind vielmehr gerade die überraschenden Befunde, die ungewöhnlichen Kombinationen von Wortmaterial, in denen sich unterschiedliche Diskurse kreuzen. Vom einzelnen Knotenpunkt ausgehend, wird sodann innerhalb eines für den Einzelfall zu

⁷³Vgl. Georg Simmel, „Der Henkel. Ein ästhetischer Versuch“, in: Ders., *Gesamtausgabe in 24 Bänden*, Bd. 7, hg. von Rüdiger Kramme u. a., Frankfurt a.M. 1995, 345–350.

definierenden Korpus aus Texten des kulturellen Archivs nach vergleichbaren Verknüpfungen – Verlinkungen – gesucht. Das konnte ich in meiner Blog-Analyse nur exemplarisch vorführen.

Tatsächlich könnte die Aufgabe des/der kulturpoetisch arbeitenden Literaturwissenschaftlers/Literaturwissenschaftlerin im Rahmen der *Digital Philology* gewissermaßen darin bestehen, eine – beispielsweise mehrere Dutzend ausgewählte, über Jahre hinweg geführte Modeblogs umfassende – Textdatenbank in einen kommentierten Hypertext zu transformieren, also etwa merkwürdige Kookkurrenzen in einem Text zu markieren sowie in einem Text erwähnte Marken mit ihren Erwähnungen in einem anderen Text zu verlinken und die sich jeweils ergebenden semiotischen Effekte zu erläutern. Die literatur- und kulturwissenschaftliche Leistung besteht so darin, die Diskurse einer Zeit aufgrund einer möglichst großen Datenbasis durch möglichst innovative *und* dem Gegenstand angemessene Verlinkungen ‚lesbar‘ zu machen. Eine Untersuchung, die literarische Phänomene im Internet auf diese Weise in ihrer ganzen textuellen Komplexität in den Blick nimmt, stellt eine adäquate Auseinandersetzung mit aktuellen literatur- und kulturwissenschaftlichen Problemen in Aussicht.

Der innovative Impuls gegenüber Morettis Methode des *Distant Reading* besteht dabei darin, dass nicht traditionelle, sondern innovative Fragestellungen oder Suchbefehle an die Texte herangetragen werden. Das setzt, nach dem Motto ‚erst finden, dann suchen‘, die exemplarische philologische Feinanalyse zumindest einzelner Texte voraus: Frage ich in zeitdiagnostischer Absicht nach Knotenpunkten von Diskursen, so muss ich sie zunächst im Kleinen finden, um in größeren Textkorpora nach ihnen suchen (lassen) zu können. Damit gilt: *Digital Philology* ersetzt keineswegs das mikroskopische philologische Handwerkszeug, aber auch: *Digital Philology* ermöglicht es der Literaturwissenschaft, sich neuen Untersuchungsfeldern wie populärkulturellen Phänomenen durch die Analyse großer Datenmengen in sinnvoller Weise zuzuwenden.

Literatur⁷⁴

Anz, Thomas, „Literaturkritik in Zeiten des Internets. Thesen zu ihrer Verteidigung und Bedenken“, in: *literaturkritik.de*, 12. Dezember 2010, http://www.literaturkritik.de/public/rezension.php?rez_id=15120.

Barthes, Roland, *Die Sprache der Mode*, Frankfurt a.M. ¹⁰2017.

Baßler, Moritz, *Die kulturpoetische Funktion und das Archiv. Eine literaturwissenschaftliche Text-Kontext-Theorie*, Tübingen 2005.

Baudelaire, Charles, „Der Maler des modernen Lebens“, in: Ders., *Sämtliche Werke und Briefe*, Bd. 5: *Aufsätze zur Literatur und Kunst 1857–1860*, hg. von Friedhelm Kemp/Claude Pichois, München/Wien 1989.

Beermann, Claire, „A Farewell Letter to Coffee. This could have been such a sweet love story“, in: *c'est clairette* (18. Juni 2017), <http://cestclairette.com/2017/06/a-farewell-letter-to-coffee/>.

⁷⁴ Sämtliche digitalen Referenzen wurden letztmalig am 24.9.2017 eingesehen.

- Beermann, Claire, „Badeanzug auf Abwegen“, in: *c'est clairette* (30. Mai 2015), <http://cestclairette.com/2015/05/badeanzug-auf-abwegen/>.
- Beermann, Claire, „Biomarkt-Falle. Verwöhnt oder anspruchsvoll“, in: *c'est clairette* (11. Juli 2016a), <http://cestclairette.com/2016/07/biomarkt-falle-verwoehnt-oder-anspruchsvoll/>.
- Beermann, Claire, „Das neue Blond. Barbie gibt's jetzt auch in klein und rund. Leider ist sie immer noch zu schön“, in: *c'est clairette* (6. Februar 2016b), <http://cestclairette.com/2016/02/das-neue-blond/>.
- Beermann, Claire, „Mode macht hungrig. Ein Monat, vier Städte, zwanzig Restaurants“, in: *c'est clairette* (11. Februar 2017), <http://cestclairette.com/2017/02/mode-macht-hungrig/>.
- Beermann, Claire, „Schlaghosen sind wie Schlagsahne. Die brandneue Klassiker-Kolumne auf C'est Clairette“, in: *c'est clairette* (20. Oktober 2014a), <http://cestclairette.com/2014/10/schlaghosen-sind-wie-schlagsahne/>.
- Beermann, Claire, „Sommer vorm Balkon“, in: *c'est clairette* (7. Juli 2014b), <http://cestclairette.com/2014/07/sommer-vorm-balkon/>.
- Beermann, Claire, „Update aus Berlin“, in: *c'est clairette* (5. Juli 2012), <http://cestclairette.com/2012/07/update-aus-berlin/>.
- Beermann, Claire, „Weisheiten 2014. Rückblick auf das Jahr des Grünkohls“, in: *c'est clairette* (30. Dezember 2014), <http://cestclairette.com/2014/12/weisheiten-2014/>.
- Dinther, Nike van, „Die Sache mit dem verlorenen Freund“, in: *This is Jane Wayne* (10. März 2013), <http://www.thisisjanewayne.com/news/2013/03/10/die-sache-mit-dem-verlorenen-freund/>.
- Dinther, Nike van, „Erste Schnapsschüsse vom Abend: Frontline Shop Fashion Show“, in: *This is Jane Wayne* (9. November 2012), <http://www.thisisjanewayne.com/news/2012/11/09/444440/>.
- Dinther, Nike van, „Film-Tipp für graue Tage & Soundtrack. ‚17 Filles‘“, in: *This is Jane Wayne* (23. Februar 2015), <http://www.thisisjanewayne.com/news/2015/02/23/film-tipp-fur-graue-tage-soundtrack-17-madchen/>.
- Dinther, Nike van, „Insta-Diary. Unsere Bilder der Woche“, in: *This is Jane Wayne* (13. Mai 2014), <http://www.thisisjanewayne.com/news/2014/05/13/insta-diary-unsere-bilder-der-woche-4/>.
- Drügh, Heinz/Metz, Christian/Weyand, Björn (Hg.), *Warenästhetik. Neue Perspektiven auf Konsum, Kultur und Kunst*, Berlin 2011.
- Görling, Kerstin, „Jourlook: Sesamstraßenkleid von Carven“, in: *Journelles* (16. September 2016), <https://www.journelles.de/jourlook-sesamstrassenkleid-von-carven/>.
- Gottschalk, Sarah, „Adventskalender. Tor 1. Der Samsung Serif TV by Ronan & Erwan Bouroullec im Wert von 1000 Euro“, in: *This is Jane Wayne* (1. Dezember 2016), <http://www.thisisjanewayne.com/news/2016/12/01/adventskalender-2016-tor-1-der-samsung-serif-tv-by-ronan-erwan-bouroullec-im-wert-von-1000-euro/>.
- Gottschalk, Sarah, „Ganz schön fruchtig“, in: *This is Jane Wayne* (29. März 2011a), <http://www.thisisjanewayne.com/news/2011/03/29/ganz-schon-fruchtig/>.
- Gottschalk, Sarah, „It's official: Jane Wayne x Viu Eyewear. Unsere limitierte Sonnenbrille ‚Arancia‘“, in: *This is Jane Wayne* (19. Juli 2017), <http://www.thisisjanewayne.com/news/2017/07/19/this-is-jane-wayne-x-viu-eyewear-unsere-limitierte-sonnenbrille-arancia/>.
- Gottschalk, Sarah, „Keine Heirat für Ernie und Bert – Das Aus der Sesamstraßen-Petition?“, in: *This is Jane Wayne* (15. August 2011b), <http://www.thisisjanewayne.com/news/2011/08/15/keine-heirat-fur-ernie-und-bert-das-aus-der-sesamstrassen-petition/>.
- Greenblatt, Stephen, *Verhandlungen mit Shakespeare. Innenansichten der englischen Renaissance*, übers. David Cackett, Berlin 1990.
- Heyden, Alexa von, „Die magische 12, das perfekte Outfit-Foto und die schönsten Klischees. So wird man laut Aimee Song („Song of Style“) zum Instagram-Star“, in: *Journelles* (16. November 2016), <https://www.journelles.de/aimee-song-song-of-style-instagram-star/>.
- Heyden, Alexa von, „Jourgarderobe: Closet Diary mit Silvia Philipp, Inhaberin einer Kölner Werbeagentur“, in: *Journelles* (1. Mai 2017), <https://www.journelles.de/jourgarderobe-close-silvia-philipp/>.
- Herbst, Alban N., „Phantastische Räume im Internet (1). Aus dem Entwurf des Marburger Vortrags“, in: *Die Dschungel. Anderswelt* (21. November 2012), <http://albannikolaiherbst.twoday.net/stories/phantastische-raeume-im-internet-1-aus-dem-entwurf-des-marburger-vortr/>.

- Jelinek, Elfriede, *Keine Anweisung. Keine Auszahlung, kein Betrag, kein Betrug*, <http://www.elfriedejelinek.com/fanmerk.htm>.
- Jürgensen, Christoph/Kaiser, Gerhard (Hg.), *Schriftstellerische Inszenierungspraktiken. Typologie und Geschichte*, Heidelberg 2011.
- Kolosowa, Wlada, „Modeblogger. Kann man denn davon leben?“, in: *DIE ZEIT* 3 (2015), <http://www.zeit.de/2015/03/modeblogger-geld-fashion-week>.
- Künzel, Christine/Schönert, Jörg (Hg.), *Autorinszenierungen. Autorschaft und literarisches Werk im Kontext der Medien*, Würzburg 2007.
- Lüders, Jenny, *Ambivalente Selbstpraktiken. Eine Foucault'sche Perspektive auf Bildungsprozesse in Weblogs*, Bielefeld 2007.
- Moretti, Franco, *Kurven, Karten, Stammbäume. Abstrakte Modelle für die Literaturgeschichte*, Frankfurt a.M. 2009.
- Moretti, Franco, *Distant Reading*, Konstanz 2016 (engl. 2013).
- Moss, Christoph/Heurich, Jill, *Weblogs und Sprache. Untersuchung von linguistischen Charakteristika in Blog-Texten*, Wiesbaden 2015.
- Nestvold, Ruth, „Das Ende des Buches. Hypertext und seine Auswirkungen auf die Literatur“, in: Martin Klepper/Ruth Mayer/Ernst-Peter Schneck (Hg.), *Hyperkultur. Zur Fiktion des Computerzeitalters*, Berlin/New York 1996, 14–30.
- O.N., „Capri? Lieber Ischia!“, in: *Süddeutsche Zeitung* (14. Mai 2010), <http://www.sueddeutsche.de/leben/eis-test-capri-lieber-ischia-1.232432>.
- O.N., „OURgarderobe: Closet Diary mit Kathrin Bommann, Produktmanagerin und Einkäuferin“, in: *Journelles* (20. März 2017), <https://www.journelles.de/jourgarderobe-closet-diary-kathrin-bommann/>.
- Ortner, Hanspeter, *Wortschatz der Mode. Das Vokabular der Modebeiträge in deutschen Modezeitschriften*, Düsseldorf 1981, <https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/1771>.
- Pedzisz, Joanna, *Profil des Online-Diskurses in Blog-Interaktionen an der Schnittstelle zwischen theoretischem Konzept und empirischem Modell*, Frankfurt a.M. 2017.
- Prüfer, Tillmann, „Modeblogs. Mode als Dialog“, in: *DIE ZEIT* 8 (2010), <http://www.zeit.de/2010/08/Modeblog>.
- Scalamari Jane [d.i. Sarah Radowitz], „10 Random Facts about ...“, in: *This is Jane Wayne* (11. April 2016a), <http://www.thisisjanewayne.com/news/2016/04/11/10-random-facts-about-scalamari-jane/>.
- Scalamari Jane [d.i. Sarah Radowitz], „Beauty. Mein Masterplan für reine Erwachsenenhaut“, in: *This is Jane Wayne* (10. Juni 2016b), <http://www.thisisjanewayne.com/news/2016/06/10/beauty-mein-masterplan-fuer-reine-erwachsenenhaut/>.
- Schildhauer, Peter, *Textsorten im Internet zwischen Wandel und Konstanz. Eine diachrone Untersuchung der Textsorte Personal Weblog*, Halle 2014, <http://digital.bibliothek.uni-halle.de/ulbhalhs/urn/urn:nbn:de:gbv:3:4-12640>.
- Schmitt, Oliver M., „Capri oder Split? Zwei Stieleis-Klassiker sind nach Badeorten in Italien und Kroatien benannt. Wo genießt es sich besser? Ein eiskalter Vergleich“, in: *DIE ZEIT* 34 (2012), <http://www.zeit.de/2012/34/Stieleis-Kroatien-Split-Italien-Capri>.
- Schuster, Jörg, „Kreativer Austausch oder digitaler Masochismus? Internet-Schreibplattformen beleben die literarische Kommunikation, wenn sie die Chancen der Interaktivität und Dezentralisation nutzen – die Beispiele literaturcafe.de und leseleupe.de“, in: *literaturkritik.de*, 4. April 2013, <https://literaturkritik.de/id/17760>.
- Schuster, Jörg, „Comeback der Philologie? In seiner nun ins Deutsche übersetzten Essaysammlung ‚Distant Reading‘ entwirft Franco Moretti eine literaturgeschichtliche Methodologie für das digitale Zeitalter“, in: *literaturkritik.de*, 9. September 2016, <https://literaturkritik.de/id/22507>.
- Simanowski, Roberto, *Interfictions. Vom Schreiben im Netz*, Frankfurt a.M. 2002.
- Simmel, Georg, „Der Henkel. Ein ästhetischer Versuch“, in: Ders., *Gesamtausgabe in 24 Bänden*, Bd. 7, hg. von Rüdiger Kramme u.a., Frankfurt a.M. 1995, 345–350.
- Thaller, Manfred, „Digital Humanities als Wissenschaft“, in: Fotis Jannidis/Hubertus Kohle/Malte Rehbein (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart/Weimar 2017, 13–18.

- Trilcke, Peer, „Ideen zu einer Literatursoziologie des Internets. Mit einer Blogotop-Analyse“, in: *Textpraxis* 7 (2013), <http://www.uni-muenster.de/Textpraxis/peer-trilcke-literatursoziologie-des-internets>.
- Venohr, Dagmar, *medium macht mode. Zur Ikonotextualität der Modezeitschrift*, Bielefeld 2010.
- Wegmann, Thomas, *Dichtung und Warenzeichen. Reklame im literarischen Feld 1850–2000*, Göttingen 2011.
- Wegmann, Thomas, „Warentest und Selbstmanagement. Literaturkritik im Web 2.0 als Teil nachbürgerlicher Wissens- und Beurteilungskulturen“, in: Matthias Beilein/Claudia Stockinger/Simone Winko (Hg.), *Kanon, Wertung und Vermittlung. Literatur in der Wissensgesellschaft*, Berlin/Boston 2012, 279–291.
- Weiß, Jessie, „Gesprächsstoff: Auch in der Mode Trend – Detox, Clean Eating und Co“, in: *Journelles* (7. Februar 2014a), <https://www.journelles.de/gesprachsstoff-auch-in-der-mode-trend-detox-clean-eating-und-co/>.
- Weiß, Jessie, „Jourgarderobe: Closet Diary“, in: *Journelles* (17. November 2014b), <https://www.journelles.de/jourgarderobe-closet-diary/>.
- Weiß, Jessie, „Jourlook: Flausch!“, in: *Journelles* (20. August 2013), <https://www.journelles.de/jourlook-flausch/>.
- Weyand, Björn, *Poetik der Marke. Konsumkultur und literarische Verfahren*, Berlin/Boston 2013.

Online-Ressourcen

- Die Dschungel. Anderswelt*, <http://albannikolaiherbst.twoday.net/>. *Fanfiction*, <https://www.fanfiction.net/>.
- Projekt Gutenberg-DE*, <https://gutenberg.spiegel.de/>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Diskussionsbericht Sektion I. Literatur(wissenschaft) unter digitalen Bedingungen

Jan Borkowski

Die drei Vorlagen, die sich mit Literaturwissenschaft unter digitalen Bedingungen befassten, setzten die etablierte Praxis aus verschiedenen Perspektiven mit Möglichkeiten in Beziehung, welche die Digitalisierung eröffnet. Sie legten dar, wie Computermodelle mit ‚traditionellen‘ literaturwissenschaftlichen Positionen, etwa historisch-hermeneutischen, in Verbindung gebracht werden können (Kuhn). Sie zeigten anhand der *Deutschen Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte*, wie eine auf der entsprechenden Auswertung großer Korpora vorgehende wissenschaftsgeschichtliche Untersuchung neue Einsichten in die fachgeschichtliche Praxis ermöglicht (Jannidis/Martus), und fragten nach den Herausforderungen, vor welche die Digitalisierung die Literaturwissenschaft stellt (Stolz). Als Gemeinsamkeit der Diskussionen erwies sich insbesondere die Behandlung damit verbundener theoretischer und methodischer Aspekte.

An der Vorlage von Michael Stolz interessierte die Teilnehmenden in erster Linie der Umgang mit bestimmten Begriffen sowie der Stellenwert des angeführten Schreibexperimentes von Lyotard, insbesondere hinsichtlich seiner Aussagekraft in Bezug auf Digitalität und ihre Folgen für die Literaturwissenschaft. Diese Punkte wurden bereits zu Beginn benannt. Zunächst sei die Frage nach dem Innovationsanspruch von Interesse: Welchen Mehrwert haben die dargestellten Überlegungen konkret, wenn es um die Beschreibung der Entwicklungen von Digitalität geht? Dazu sei auch zu bemerken, dass die in der Vorlage einleitend formulierte grundlegende These zu einem medialen Wandel in den weiteren Ausführungen, anders als erwartet, nicht strukturbildend sei. Die Implikationen des eingehender beschriebenen Schreibexperimentes könnten expliziter gemacht werden. Des Weiteren sei es erforderlich, genauer zu

J. Borkowski (✉)
Universität Göttingen, Göttingen, Deutschland
E-Mail: Jan.Borkowski@phil.uni-goettingen.de

bestimmen, was hier ‚Literatur‘ und ‚Literaturwissenschaft‘, ‚der Autor‘ und ‚der Leser‘ heißen sollen. Ferner wurde gefragt, in welchem Verhältnis die durch die Vorlage eröffneten Potenziale zu empirischen Studien stünden.

Eine erste Bemerkung zur Terminologie betraf den Begriff ‚Lesen‘. Es wurde beobachtet, dass damit in der Vorlage Unterschiedliches gemeint sei: das Entziffern von Text, die kognitive Aufnahme von Informationen, das Zuweisen von Bedeutung und Lesen als Erfahrung. Zweitens wurde der Überzeugung Ausdruck verliehen, dass ‚Literatur‘ und ‚Literaturwissenschaft‘ sich einander aufgrund der auch in der Vorlage beschriebenen gegenwärtigen Entwicklungen der Digitalisierung annäherten. In einer anderen Wortmeldung wurde diese Aussage allerdings bestritten. ‚Literatur‘ bezeichne weiterhin den Gegenstand und ‚Literaturwissenschaft‘ ein bestimmtes Sprechen über eben diesen Gegenstand. Damit gebe es nach wie vor eine klare Trennung. Drittens wurde insistiert, dass die Verwendung von Kollektivsingularen in der Vorlage Probleme mit sich bringe. ‚Die‘ Literaturwissenschaft gebe es nicht, weder fachgeschichtlich noch zeitgenössisch. Es hätten stets unterschiedliche, konkurrierende Literaturwissenschaften nebeneinander bestanden. Ebenso wenig könne man einfach von ‚dem‘ Lesen sprechen, da in der Geschichte und auch heute stets verschiedene Lesekulturen existierten.

Was das dargestellte Schreibexperiment und seinen Aussagewert betrifft, wurde zunächst um eine Klärung gebeten, ob hier Schlussfolgerungen gezogen würden, die ungültig seien. Sei es nicht, so eine erste Nachfrage, ein Fehlschluss, wenn man von Positionen, wie sie im Rahmen des Poststrukturalismus formuliert wurden, auf Entwicklungen im Bereich des Digitalen schließe? Müssten entsprechende Aussagen anders validiert werden? Eine zweite Nachfrage ging in die gleiche Richtung: Liege nicht ein Zirkelschluss vor, wenn man von postmodernen Theorien auf medienwissenschaftliche Annahmen schließe? Bei einer dritten Nachfrage ging es darum, ob bei Feststellungen über Prozesse des Wandels in der Vergangenheit die Gefahr einer Fehleinschätzung bestehe, weil sie aus heutiger Sicht anders – beispielsweise eindeutiger – erscheinen, als sie es tatsächlich oder auch in der Wahrnehmung der Zeitgenossen waren. In einem Diskussionsbeitrag wurde vermutet, dass sich an dem vorgestellten Schreibexperiment vor allem ein zeitgenössischer Medienwandel beobachten lasse. In einem weiteren Diskussionsbeitrag wurde um eine genauere Beschreibung des gegenwärtigen Zustandes gebeten, für den in der Vorlage auf Heideggers Begriff der ‚Zuhandenheit‘ zurückgegriffen wurde. Ferner wurde gefragt, ob der Fokus auf Synchronie in der Vorlage eine Skepsis gegenüber der Möglichkeit von Innovationen zum Ausdruck bringe. Was die verschiedenen Literaturtheorien betrifft, die es im Fach gibt, wurde beobachtet, dass sie in unterschiedlicher Weise geeignet seien, im Rahmen einer an Phänomenen des Digitalen interessierten Literaturwissenschaft operationalisiert zu werden. Es gebe solche, bei denen der Grad der Operationalisierbarkeit hoch, und solche, bei denen er niedrig sei. Angeregt wurde, zeitgenössische Schreibexperimente aus dem Bereich des Populären, etwa Spielarten von *Textadventures* im Internet, in die Betrachtung einzubeziehen. Das

könne sich als fruchtbare Ergänzung zu dem untersuchten avantgardistischen Schreibexperiment erweisen.

Stolz verwies auf den heuristischen Wert der von ihm vorgestellten Überlegungen. An dem besagten Schreibexperiment würden grundlegende Besonderheiten deutlich, die den medialen Wandel und den Umgang mit diesem betreffen. Dies lasse sich prinzipiell auch auf andere Zusammenhänge wie aktuelle Tendenzen der Digitalisierung übertragen. Es sei nicht beabsichtigt gewesen, bestimmte Begriffe ausschließlich im Kollektivsingular zu verwenden. Die zu beobachtenden, durch die Digitalisierung bewirkten tiefgreifenden Wandlungsprozesse brächten terminologische Unschärfen und einen entsprechenden Klärungsbedarf mit sich; manche Begriffe, darunter ‚Lesen‘, seien unter diesen Bedingungen neu zu fassen. Für Phänomene, die sich im Zuge der Digitalisierung erst allmählich abzeichnen, stehe aber (wie bei jedem kulturellen Wandel) vorab nur die zu diesem Zeitpunkt vorhandene Beschreibungssprache zur Verfügung.

Jonas Kuhns Vorlage regte insbesondere methodische Überlegungen an und konkrete verfahrenstechnische Fragen zur vorgestellten Untersuchung. Als zentraler Punkt wurde das Verhältnis quantitativer und qualitativer Methoden identifiziert, vor allem ihre mögliche Kompatibilität in bestimmten Hinsichten: Wie genau lassen sich quantitative Verfahren, die in der Regel weniger komplexe Phänomene auf ‚unteren‘ Ebenen mithilfe von Annotationen erfassen, mit traditionellen, etwa hermeneutischen Verfahren literaturwissenschaftlicher Forschung in Verbindung bringen, die sich üblicherweise auf ‚höherer‘ Ebene bewegen und komplexere Phänomene in den Blick nehmen?

Methodische Überlegungen bezogen sich auf die Potenziale von Annotationen, auf das Verhältnis des vorgestellten Ansatzes zu literaturwissenschaftlichen Methoden und auf die Korpuswahl. Es wurde vorgeschlagen, dass ein größerer Optimismus mit Blick auf die Möglichkeit geboten sei, Konsensannotationen zu erstellen. Eine Differenzierung sei erforderlich, weil es mehr Fälle als üblicherweise angenommen gebe, in denen es eher einfach sei, Konsens bei der Annotation zu erzielen. Das gelte jedenfalls für weniger interpretationsbedürftige literarische Texte bestimmter Gattungen und Epochen. Sie seien von komplexeren Fällen abzugrenzen, bei denen es sich tatsächlich als schwieriger erweisen könne, konsensuelle Annotationen vorzunehmen. Zudem sei zu beachten, dass selbst bei der Annotation von Phänomenen auf ‚höherer‘ Ebene (als Beispiel wurde die Argumentationsanalyse genannt) in der Praxis die Chance auf Konsens durchaus größer sei, als man auf den ersten Blick vielleicht meinen möchte. In die gleiche Richtung ging die Frage, welche Rolle weitgehend konsensuelle Entscheidungen spielen können, die sich im Rahmen der Textanalyse ergeben. Sie seien als großer Gewinn zu betrachten und würden in der Vorlage vielleicht in ihrer Rolle und Relevanz etwas unterschätzt. Mehr Konvergenz bei der Textanalyse sei sehr erstrebenswert, sie sei ein wichtiger Beitrag zur Verwissenschaftlichung und Empirisierung der Literaturwissenschaft.

Was das Verhältnis des vorgestellten Ansatzes zu literaturwissenschaftlichen Methoden angeht, wurde um eine Klarstellung gebeten, ob beabsichtigt sei, hermeneutischen Zugangsweisen zuzuarbeiten oder eine Alternative zu

ihnen anzubieten. Sollte Letzteres der Fall sein, sei zu beachten, dass es in der Literaturwissenschaft nicht nur hermeneutische Positionen gebe, sondern auch andere, zum Beispiel strukturalistische. Daran thematisch anknüpfend, wurden in zwei Wortmeldungen Präzisierungen vorgeschlagen. Die verschiedenen literaturtheoretischen Ansätze wiesen eine unterschiedlich gelagerte Passung mit und Anschlussfähigkeit für Methoden und Verfahren computerlinguistischer Art auf. Sie eigneten sich daher unterschiedlich gut für eine Formalisierung. Das Ziel, hermeneutische Positionen zu formalisieren und zu objektivieren, sei als besonders fruchtbar einzuschätzen. Diese Bewertung wurde in einer weiteren Wortmeldung bekräftigt: Der vorgestellte Ansatz verfüge über große Potenziale, zumindest für den Teil des Faches, der an der Operationalisierung und Präzisierung von Verfahrensweisen Interesse habe.

Im Hinblick auf die Wahl des Korpus wurde auf mögliche Unterschiede zwischen (Computer-)Linguistik und Literaturwissenschaft beziehungsweise *Computational Literary Studies* aufmerksam gemacht. Untersuche die Computerlinguistik nicht, so die Ansicht in einem Diskussionsbeitrag, Sprache als System, während es den *Computational Literary Studies* um die Analyse historischer Ereignisse gehe? Könne es sein, so ein anderer Diskussionsbeitrag, dass beim bisherigen Stand die Arbeit mit speziell für bestimmte Fragestellungen erhobenen Korpora in der Literaturwissenschaft die einzige sinnvolle Möglichkeit darstelle? Sei es außerdem denkbar, ein Korpus auszuwerten, das Artefaktstatus besitzt, zum Beispiel im Falle von vorhandenen historischen Sammlungen, die eigentlich mit anderen Absichten zusammengestellt worden waren? Oder handle es sich um Fälle, die zugunsten eines anvisierten Goldstandards überwunden werden müssten? In diesem Zusammenhang wurde auch die Frage aufgeworfen, wann eine Grundgesamtheit als groß genug einzuschätzen sei, um statistische Verfahren, etwa der Inferenzstatistik, einzusetzen. Erwogen wurde des Weiteren, ob einem bei der Überprüfung von Hypothesen im Zusammenspiel von konzeptionellen Vorstellungen, eingesetzten Verfahren und gewähltem Korpus ein Zirkelschluss unterlaufen könne.

Konkrete verfahrenstechnische Hinweise setzten bei dem behandelten Beispiel an. Die Wahl einer narratologischen Kategorie – Fokalisierung in ausgewählten Erzähltexten Arthur Schnitzlers – könne sich, so wurde angemerkt, als unglücklich erweisen mit Blick auf das, was gezeigt werden solle. In diese Richtung ging auch die Vermutung, dass bei einer solchen Wahl die Gefahr bestehen könne, leichter formalisierbare Interpretationshypothesen zu privilegieren. Zur Arbeit mit Übersetzungen wurde gefragt, ob sie nicht ein methodisches Problem darstelle, da es verschiedene Möglichkeiten gebe, einen literarischen Text zu übersetzen, hinter denen unterschiedliche Konzeptionen von Übersetzung stehen. Erwogen wurde schließlich auch, ob die Arbeit mit Leseindrücken bei der Annotation zur Konsequenz habe, dass manche relevanten Differenzen nicht erfasst werden können.

Kuhn stimmte der Einschätzung zu, dass Konsensannotationen in den genannten Bereichen möglich und sinnvoll seien, betonte die Kompatibilität der vorgestellten Verfahren mit verschiedenen literaturtheoretischen Ansätzen,

bekräftigte den Aussagewert computerlinguistischer Korpora und pflichtete der Einschätzung bei, der zufolge in bestimmten Fällen auch andere Formen der Bildung von Korpora sinnvoll seien. Außerdem nahm er die Hinweise zum Anlass, verfahrenstechnische Details näher zu erläutern.

Bei der Diskussion der Vorlage von Fotis Jannidis und Steffen Martus standen Bemerkungen zur Durchführung der Experimente und zu ihren Ergebnissen sowie Anregungen und Überlegungen zu weiterführenden Studien im Mittelpunkt. Positiv hervorgehoben wurde zunächst die Entscheidung, in der Vorlage auch Experimente zu erwähnen, die ohne Befund geblieben sind. Angeregt wurde zum einen, bei der Erklärung mancher Befunde (z. B. der quantitativen Sprünge beim Fußnotenvorkommen im Untersuchungskorpus) die zu der jeweiligen Zeit vorhandenen Bibliotheksbestände und ihre Verfügbarkeit als Faktor einzubeziehen und nicht allein wissenschaftsgeschichtlich zu argumentieren: So seien etwa erst in den 1960er Jahren die kriegsbedingten Verluste kompensiert gewesen, während in den 1990er Jahren die Umstellung auf EDV erfolgt sei. Zum anderen wurde beobachtet, dass in der Vorlage Erklärungen vor allem auf die Intentionen der Akteure Bezug nehmen, und vorgeschlagen, demgegenüber Erklärungen zu bevorzugen, die ‚im Rücken‘ der Akteure anzusiedeln sind. Kritisiert wurde, dass die Annahmen über die *DVjs*, welche die Wahl des Korpus legitimieren sollen, nicht hinreichend belegt worden seien. Manche Ergebnisse seien erwartbar, etwa der Befund, dass die gewählte Zeitschrift mit vergleichbaren in engerer Verbindung stehe, nicht jedoch mit zum Beispiel fachdidaktischen.

Es wurde die Frage aufgeworfen, wie sich ertragreiche Experimente anlegen lassen. Aus der Verwendung von Wörtern, die auf ‚-ung‘ und ‚-bar‘ enden, ließen sich, um ein Beispiel zu geben, anders als in der Vorlage angenommen, Aussagen über den Gebrauch einer wissenschaftlichen Terminologie eventuell nicht in idealer Weise ableiten. Zudem wurde angeregt, die Erschließung sprachlicher Quellen in der Linguistik zum Vorbild zu nehmen. Auf diese Weise könne unter anderem die Gefahr eines *Confirmation Bias* verringert werden. Auch stelle sich die Frage, wie sich die Ergebnisse solcher Untersuchungen angemessen visualisieren lassen. In der Vorlage sei dies nicht immer optimal gelöst worden.

Weitere Bemerkungen zur Durchführung der Experimente und zu ihren Ergebnissen setzten an verschiedenen Stellen an. Bestehe nicht, so wurde gefragt, die Notwendigkeit, den zugrunde liegenden praxeologischen Ansatz und die mit ihm einhergehenden Vorannahmen transparenter zu machen, da ein solcher Ansatz zwar manche Praktiken erfasse, andere, etwa ‚harte‘ Praktiken der Wissenschaftsgeschichte, jedoch nicht? In einer anderen Wortmeldung wurde der Standpunkt der Verfasser der Vorlage thematisiert: Müssten sie sich nicht positionieren, da sie hier einen Beobachterstandpunkt einnehmen, aber zugleich auch Vertreter des Faches sind und an dessen Praktiken partizipieren? Ein weiterer Diskussionspunkt betraf die Möglichkeit, den gewählten quantifizierenden Ansatz mit qualitativen Analysen zu verbinden bzw. die auf diesem Wege gewonnenen Ergebnisse mit zum Beispiel hermeneutischen Erklärungen. Auf Skepsis stieß in einem Beitrag die Erklärung mancher Befunde unter Hinweis auf Bestrebungen der Verwissenschaftlichung in der Fachgeschichte. Dazu wurde die Zahl der Publikationen

zu diesem Thema ins Spiel gebracht, die etwa in den 1970er Jahren eher klein gewesen sei. In einer Wortmeldung wurde noch einmal die Rolle angesprochen, welche die Bibliotheksinfrastruktur für die Erklärung mancher Befunde spielen könne: Gebe es zum Beispiel eine Korrelation zwischen der Menge an verfügbarer (aktueller) Forschungsliteratur und der Anzahl der Fußnoten in den Aufsätzen der *DVjs*? Eine weitere Nachfrage betraf den Stellenwert bestimmter Gegebenheiten im Untersuchungskorpus, etwa das Vorliegen von Sonderheften oder den unterschiedlichen Umfang der Hefte, der unter Umständen bestimmten Mustern folge (z. B. nach einem längeren stets ein kürzeres Heft).

Zu den Anregungen und Überlegungen für weiterführende Studien gehörte die Frage, ob die gewählten Indikatoren spezifiziert werden können. Bei Fußnoten zum Beispiel ließen sich aussagekräftige weiterführende Ergebnisse erzielen, wenn ihr Inhalt (bibliographische Hinweise, Nebendiskussionen u. a.) einbezogen würde. Es wurde eine Reihe von Vorschlägen gemacht. Man könne Netzwerkanalysen vornehmen, bei welchen nicht die Autoren, sondern die Zeitschriftenbeiträge die Knoten bilden; bei den Experimenten, die keine Ergebnisse erbracht haben, das Korpus ausweiten; stilistische Originalität in dem gewählten Korpus anhand einschlägiger Kriterien (Wortarten, *type/token-ratio*, Abweichungen mit Blick auf das Verhältnis des Themas zur Wortwahl usw.) und mithilfe statistischer Verfahren ermitteln; eine Typologie der Zeitschriftenartikel hinsichtlich ihrer Struktur erstellen; über die bisher erfassten Strukturinformationen – wie Titel, Untertitel, Ort der Beiträge, Zitate und Fußnoten – weitere für *TEI* aufbereiten.

Martus und Jannidis bewerteten die Hinweise zu den durchgeführten Experimenten und die Vorschläge für zukünftige Studien in der Mehrzahl als wichtig und sinnvoll, mitunter seien sie jedoch mit methodischen Schwierigkeiten verbunden oder sie beruhten auf unzutreffenden Annahmen. Der praxeologische Ansatz sei hier hinreichend leistungsfähig, Fragen der Standortgebundenheit stellten sich nicht in der angenommenen Form. Sie pflichteten der Ansicht bei, dass bibliotheksgeschichtliche Faktoren eine wichtige Rolle spielen könnten bei der Erklärung einiger Befunde, verwahrten sich allerdings gegen manche der vorgebrachten Kritikpunkte. Die Wahl von Aufsätzen aus der *DVjs* sei durchaus nicht willkürlich, die in der Vorlage genannte Begründung hinreichend. Das Korpus habe einen hohen Aussagewert. Die Wahl von Wörtern, die auf ‚-ung‘ und ‚-bar‘ enden, für Rückschlüsse auf den Einsatz von wissenschaftssprachlichen Ausdrücken bediene sich eines objektiven und operationalisierbaren Kriteriums, das aus der Linguistik bekannt sei.

Die drei Vorlagen, die Literatur unter digitalen Bedingungen behandelten, gaben anhand ausgewählter Beispiele Einblicke in die unterschiedlichen Formate und Praktiken des Umgangs mit Literatur im digitalen Bereich. Im Einzelnen ging es um populäre Philologie im Internet am Beispiel von *genius.com* (Nebrig), die Produktion und Rezeption literarischer Texte unter digitalen Bedingungen am Beispiel von *Tausend Tode schreiben* und *Ox0a* (Nantke) sowie um die Untersuchung von Mode-Blogs (Schuster). Als Gemeinsamkeit sowohl der Vorlagen als auch der Diskussionen erwies sich das Bestreben, methodisch abgesichert und

konzeptionell reflektiert das jeweils mit Blick auf das digitale Medium Spezifische dieser Artefakte und der zugehörigen kommunikativen Praktiken zu erkennen.

Die Vorlage von Alexander Nebrig wurde vor allem mit Blick auf mögliche Präzisierungen ihrer Befunde und ein vertiefendes Verständnis des behandelten Phänomens betrachtet. Am Beginn stand eine Einschätzung der Aussage aus der Vorlage, wonach bei dem in Rede stehenden Sachverhalt – Erschließung und Kommentierung popkultureller Lieder (vor allem Rap-Texte) auf der Internetplattform *genius.com* – davon gesprochen werden könne, dass hier netzbedingte Praktiken literarisiert würden. Diese Aussage sei zu stark und sachlich nicht ganz richtig, da es die entsprechenden Praktiken bereits vorher gegeben habe. Nicht der Gegenstand sei durch das Internet verändert worden, sondern der Umgang mit ihm, das heißt die (laien-)philologische Praxis. Zu erwägen sei auch, dass es in der Literaturwissenschaft eine Liebhaber-Tradition gebe, nicht erst bei den in der Vorlage untersuchten Formen des Umgangs mit Texten. Zudem wurde gefragt, ob die für die besagte Wissensdatenbank kennzeichnende kollektive Praxis außerwissenschaftlicher Experten nicht überinterpretiert werde, wenn man sie als ‚wild‘ bezeichne. Es handele sich immerhin um eine betriebswirtschaftlich organisierte Plattform, die auch durch interne Kontrollmechanismen und finanzielle Interessen geprägt sei.

Nicht ganz einleuchten wollte in einem Diskussionsbeitrag das Vorgehen, den Begriff der Laienphilologie über die Akteure zu bestimmen (Amateure statt professioneller Akteure). Seien nicht vielmehr die Plattform und der Ort ausschlaggebend? Es wurde die Einschätzung vertreten, wonach Rap-Texte ein derart spezifischer Gegenstand seien, dass sie andere Zugangsweisen erforderten als philologische; der Gegenstand beeinflusse hier in starkem Maße die zu beobachtenden Praktiken. Nachgefragt wurde des Weiteren, als wie hoch der Grad an Professionalisierung auf dieser Plattform eigentlich einzuschätzen sei. Es sei zu bemerken, dass das Gros der dort zu findenden Beiträge weniger professionell sei und zum Beispiel triviale Aussagen enthalte oder das Resultat identifikatorischer Lektüremuster sei. Angemerkt wurde dazu in einem anderen Redebeitrag, dass es zu Rap-Texten auf der Plattform sehr gute Stellenkommentare gebe und ‚Laienphilologie‘ hier hinsichtlich der Art und des Umfangs der bereitgestellten Informationen zu bestimmen sei.

Es wurde beobachtet, dass die Beschreibung der Plattform in der Vorlage vielleicht nicht ausführlich genug sei, und vorgeschlagen, dies noch nachzutragen. Die aufgestellten Thesen seien nicht in dem wünschenswerten Maße durch die angeführten Beobachtungen gedeckt. Gefragt wurde auch, welche Funktionen die Annotationen für welche Leser beziehungsweise Nutzer hätten, ob sie zum Beispiel dem Textverständnis dienten und welcher Wissensbegriff zugrunde liege. Außerdem wurde nach Kriterien für die Annotationen gefragt und darauf hingewiesen, dass der praktische Umgang mit ihnen mitunter Probleme mit sich bringe, etwa wenn einmal vorgenommene Annotationen zu einem späteren Zeitpunkt gelöscht würden und dann nicht mehr auffindbar seien. Angemerkt wurde auch, dass gegenüber den Referenzbeispielen (Jacob Grimm, Herder) eine stärkere Differenzierung geboten sei hinsichtlich Überlieferung und Kommentar.

Schließlich wurde zum einen die Frage nach der Historizität der beobachteten Praktiken gestellt. Schreiben, Lesen, Edieren, Kommentieren und dergleichen seien ubiquitäre Praktiken; als entscheidend erweise sich, wann sie in welcher Kombination auftreten und damit eine Praxis konstituierten. Es sei anzuregen, darauf hin die Plattform genauer in den Blick zu nehmen, andernfalls sei die Untersuchung in dieser Hinsicht nicht vollständig. Zum anderen wurde nahegelegt, Hierarchien, Disziplinierungsmechanismen, Ausschlussverfahren und ihre Funktion, Verfahren der Qualitätserzeugung und dergleichen genauer zu betrachten.

Nebbrig präziserte, dass ‚Wildheit‘ in der Vorlage relational zu verstehen sei, also das Ergebnis eines Vergleichs der untersuchten Praxis mit anderen Praktiken darstelle. Zu Recht werde auf gegenläufige Mechanismen hingewiesen, zu bedenken sei jedoch insbesondere die Nutzerebene: Dort entstehe der Impuls zur Teilnahme aus Liebhaberei, ökonomische Interessen seien demgegenüber nachgeordnet. Die Praxis wandele sich durch das Internet; anzunehmen sei tatsächlich nicht, dass sie erst durch das Internet entstehe. Neuartig seien die Partizipationskultur und die Beteiligung der Rezipienten beziehungsweise Nutzer an der Produktion, ferner Art und Umfang, in dem es möglich sei, solche Gegenstände textkritisch zu behandeln und zu beurteilen. Die Wichtigkeit der Akteure sei zu betonen. Der unterstellte Zusammenhang zwischen Gegenstand und außerphilologischen Praktiken bestehe in dieser Form wohl nicht. Es gebe auf der Plattform unterschiedliche Liebhabergemeinden mit unterschiedlichen Professionalisierungsniveaus. Zu beachten sei, dass es dort neben der Kommentierung auch um das Edieren von Texten gehe, welches als besondere Leistung anzusehen sei. Man könne zwar einzelne Nutzerkommentare als identifikatorisch oder naiv bezeichnen. Wohl kaum wird man aber die quantitative Dimension des durch Verschriftung entstandenen Textkorpus leugnen. In der Vorlage hätten Systematisierungsbemühungen den Vorrang gehabt vor Beschreibungen der Befunde.

Julia Nantkes Vorlage wurde vor allem betrachtet hinsichtlich der Klassifikation, begrifflichen Beschreibung und Einschätzung der behandelten Beispiele digitaler Literatur. Positiv hervorzuheben seien, wie einleitend festgestellt wurde, der Gegenstandsbezug sowie die Vielfalt der untersuchten Phänomene und die daraus gewonnenen Einsichten. Außerdem handle es sich um gute Beispiele. Es werde deutlich, dass es zwischen den einschlägigen Parametern Verschiebungen gebe, dies könne über den Fokus auf die Kategorie ‚Text‘ hinaus stärker gemacht werden. Relevante Einsichten ergäben sich auch in die Kategorie ‚Autorschaft‘, da in der Vorlage unter anderem Autorschaft an den Computerprogrammen und unfreiwillige Autorschaft eine Rolle spielten. Zu prüfen sei, wie sich ‚Autorschaft‘ hier im Einzelnen fassen lasse; dazu seien auch Fragen der Urheberschaft zu bedenken. Außerdem könne das Verhältnis von professionellen Akteuren und Amateuren eingehender beschrieben werden. Eine Differenzierung sei erforderlich hinsichtlich der verschiedenen Formen von Literatur im Internet, etwa digitaler Literatur, Netzliteratur und Maschinenpoesie.

Es wurde die Frage aufgeworfen, was an den analysierten Texten eigentlich spezifisch sei für Digitalität. Bei manchen traditionellen Anthologien könne

Vergleichbares beobachtet werden: Auch hier gebe es verschiedene Autoren und wirkten unterschiedliche Texte zusammen. In einem weiteren Diskussionsbeitrag wurde auf populäre Anthologie-Projekte hingewiesen. Es wurde nahegelegt, dass für die in Rede stehenden Gegenstände der Werkbegriff angemessener sei als der Textbegriff, und vermutet, dass den analysierten Texten ein starker Werkbegriff zugrunde liege. Außerdem wurde gefragt, ob man in Anbetracht von Phänomenen wie den beschriebenen den Textbegriff neu bestimmen müsse. Zu klären sei dann: Wie genau sähe eine solche Neubestimmung aus? Was sei zum Beispiel mit Merkmalen wie dem Vorliegen ‚bedeutungsstiftender Relationen‘? Inwieweit seien Bedeutungszuweisungen bei Beispielen wie den behandelten überhaupt von Belang? Außerdem wurde die These aufgestellt, der zufolge das Spezifische der Digitalität in solchen Fällen vor allem bei der Distribution und Rezeption zur Geltung komme, nicht jedoch bei Autor und Werk. Hierher gehörte auch die in einem weiteren Diskussionsbeitrag gestellte Frage, ob es sich bei dem Untersuchten um Konzeptkunst handle, die keine Bedeutungsattribution benötige, weil die Idee im Mittelpunkt stehe, nicht der Text, und es um das Ausstellen einer künstlerischen Konzeption gehe. Dazu wurde in einem weiteren Redebeitrag bemerkt, dass beim Umgang mit Texten dieser Art auch ästhetische Erfahrung relevant sei, nicht allein die Wahrnehmung als Konzeptkunst. Zudem seien mitunter bestimmte Funktionen denkbar, es könne zum Beispiel Kritik an politischen Verhältnissen zum Ausdruck gebracht werden.

Vermutet wurde, dass Phänomene wie die untersuchten sich weitgehend mit traditionellen Verfahren beschreiben ließen. Dagegen wurde jedoch von anderer Seite betont, dass die Form der Distribution neu sei und spezifisch für Digitalität. Es seien neue Verfahren erforderlich, um diese Phänomene zu beschreiben; das könne nicht allein mit bewährten Mitteln erfolgen, da andere Faktoren zu beachten seien. Schließlich wurde die Einschätzung mitgeteilt, wonach in der Vorlage mitunter zu starke Thesen aufgestellt würden, etwa bezüglich des Projekts *Durchschnitt*, und es wurde der Überzeugung Ausdruck verliehen, dass die untersuchten Beispiele für digitale Literatur stärker in den größeren mediengeschichtlichen Zusammenhängen zu betrachten seien, in welchen sie stehen, um so durch Vergleich möglichen Fehleinschätzungen vorzubeugen.

Nantke erläuterte, dass die Kategorie ‚Text‘ als konzeptioneller Ausgangspunkt gewählt worden sei. Bei den Beispielen für digitale Literatur seien verschiedene Autorfunktionen und Formen der Autorschaft (kollaboratives, kooperatives Schreiben usw.) von Belang. Die untersuchten Formate seien nicht als gänzlich neu einzuschätzen, ihre Beschaffenheit und der Umgang mit ihnen sei jedoch durch Digitalität signifikant geprägt. Verglichen mit Anthologien hätten sich die entsprechenden Möglichkeiten bei digitaler Literatur stärker etabliert, Art und Umfang der Bezüge zwischen den Texten seien spezifisch. Es gebe hier auch einen engen Zusammenhang zwischen Text- und Werkbegriff. Der Textbegriff sei in der Tat neu zu bestimmen. Texte dieser Art würden nicht in Gänze rezipiert, allerdings in Ausschnitten durchaus einer ‚dichten‘ Lektüre unterzogen. Dabei werde auch eine Bedeutung zugewiesen. Die Rezeption sei nicht auf das Erkennen der künstlerischen Konzeption zu beschränken.

Die Vorlage von Jörg Schuster gab vor allem Anlass zu methodischen Überlegungen, auch unter Berücksichtigung der Besonderheiten des Gegenstandes. Am Anfang stand der Ergänzungsvorschlag, die Kategorien ‚Autor‘ und ‚Werk‘ stärker in die Untersuchung einzubeziehen; man könne den Eindruck haben, dass die in der Vorlage behandelten Texte nicht mehr als Produkte von Menschen wahrgenommen würden, sondern als maschinengenerierte Texte. Hinzuweisen sei auf den wichtigen Umstand, dass Korpora wie die ausgewerteten aufgrund der Digitalität stets von allen einsehbar seien; dies sei etwas Neues. Die analysierten Texte böten zudem neue Möglichkeiten, Autorschaft und ihre Genese zu beobachten.

Eine erste methodische Frage bezog sich darauf, welche Eigenschaften und Funktionen die in der Vorlage ausgemachten ‚Knotenpunkte von Diskursen‘ aufweisen, wie sie zu identifizieren und von Zufallsfunden zu unterscheiden seien. Dazu wurde angeregt, systematisch vorzugehen, indem man alles das, was zur selben Kategorie wie der Zufallsfund gehöre, auf mögliche Befunde prüfe. Ferner wurde darum gebeten, genauer darzulegen, wie sich die geschilderten ‚Irritationen‘ beschreiben ließen. Das Irritationspotenzial der in der Vorlage beschriebenen Fälle sei am Ende eventuell nicht ganz so groß wie angenommen. Es wurde vorgeschlagen, auf den Begriff des ‚Interdiskurses‘ zurückzugreifen. In einem Diskussionsbeitrag wurde nahegelegt, eine soziologische oder kulturgeschichtliche Untersuchung vorzunehmen; ein solcher Zugang sei besser geeignet als ein literaturwissenschaftlicher.

In einem Redebeitrag wurde kritisiert, dass in der Vorlage *Distant Reading* in unzulässiger Weise mit ‚Suchen‘ gleichgesetzt werde. Das Vorgestellte sei aber anschlussfähig an und geeignet für *Distant Reading*, auch hinsichtlich der Funktion einer Vorfilterung der einzubeziehenden Texte. Es wurde die Frage aufgeworfen, wie damit umgegangen werden solle, dass im Internet veröffentlichte Dokumente potenziell instabil seien. Angeregt wurde in diesem Zusammenhang auch, netzwerktheoretische Lösungen in die Untersuchung einzubeziehen; so könne die vorgestellte Verfahrensweise reproduzierbar gemacht werden. Wie damit umzugehen sei, dass das gewählte Verfahren es notwendigerweise mit sich bringe, Texte aus ihren Verwendungszusammenhängen im Internet herauszunehmen, war Gegenstand einer weiteren kritischen Frage. In einem anderen Redebeitrag wurde der methodische Vorbehalt, wonach Texte dieser Art nicht auch dann ausgewertet werden könnten, wenn man sie herunterlädt und ‚offline‘ behandelt, als unbegründet angesehen. Es wurde dazu ermuntert, deutlicher herauszustellen, was genau computerphilologische Verfahren für das in der Vorlage Thematisierte leisten können.

Gefragt wurde auch, worin die Spezifika der besagten Blogs bestehen. Der in der Vorlage zugrunde gelegte weite Literaturbegriff müsse genauer bestimmt werden. Zu beachten sei, dass diese Blogs in starkem Maße mit Bildern arbeiteten, die eventuell eine medienwissenschaftliche Herangehensweise erforderlich machten. Es wurde dafür plädiert, an dem verwendeten ‚kulturpoetischen‘ Ansatz konsequent festzuhalten, der zwar in computerphilologischer Hinsicht nicht überzeugen könne, mit seinem Fokus auf ‚Textualität‘ allerdings als konzeptionell fruchtbar einzuschätzen sei.

Schuster hob unter anderem hervor, dass ein ‚mikroskopischer‘ Zugang gewählt worden sei, bei dem auf induktivem Wege die konkreten Fragestellungen generiert wurden („erst finden, dann suchen“); erst auf dieser Grundlage ergebe der Einsatz maschineller Verfahren Sinn. Zufallsfunde beim Lesen, die ungewöhnlich seien und damit ‚Irritationen‘ bewirkten, bildeten den Ausgangspunkt für entsprechende Suchen mithilfe maschineller Verfahren. Es sei das Zusammentreffen zweier Diskurse in den behandelten Beispielen, welches die besagte Irritation bewirke. Auch wenn der untersuchte Gegenstand unter einen weiten Literaturbegriff falle, seien für die ‚mikroskopische‘ Analyse von Texten genuin literaturwissenschaftliche Verfahren erforderlich.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.



Sektion II. Digitale Edition und Annotation



Einführung

Andrea Rapp

Digitale Editionen und Annotationen wurden in dieser Sektion als Ergebnis wie auch als Grundlage und Ausgangspunkt wissenschaftlicher Forschung betrachtet. Die Perspektiven auf Edition umfassten sowohl den digitalen Produktionsprozess der Erstellung von Editionen als auch die erweiterten Rezeptions- und Auswertungsmöglichkeiten digitaler Präsentationsformen. Vielfältige Bezugspunkte ergaben sich darüber hinaus zu den Themen der anderen Sektionen, wenn beispielsweise digitale Editionen und annotierte Korpora als Grundlage digitaler Analyseverfahren adressiert wurden oder die Rolle von Bibliotheken, Archiven und Verlagen bei der Erstellung und Bewahrung digitaler Editionen bedacht wurde. In dieser Sektion gab es zumindest zwei Autorenkollektive, was vielleicht als symptomatisch für den Gegenstand genommen werden kann, sind doch Edition häufig Teamleistungen und Annotationen zentrale Werkzeuge kollaborativen Erschließens und Analysierens. Insbesondere die Diskussionen zeigten, dass einige Kernthemen des Kolloquiums aufgegriffen und fortgesetzt bzw. skizziert werden konnten, dass aber zugleich dieses Thema seinen Eigenwert hat und die Editionsphilologie eigene spezifische Diskurse führt. Edieren und Annotieren sind grundlegende Aktivitäten philologischer Arbeit, die deutliche Berührungspunkte und Überschneidungen aufweisen und daher in diesen Überschneidungsbereichen durchaus mit ihren Querverbindungen und Wechselwirkungen zusammen betrachtet werden können. Dazu gehören z. B. Auseinandersetzungen mit der Rolle des Kommentars, wobei dieser Blick noch geweitet wird, wenn das Verhältnis von Kommentar und literaturwissenschaftlicher Interpretation angesprochen ist. Eine getrennte Betrachtung ihrer Eigenlogiken ist legitim, um Standortbestimmungen, Reformulierungen und Evolution der Tätigkeiten des Edierens und Annotierens

A. Rapp (✉)
Technische Universität Darmstadt, Darmstadt, Deutschland
E-Mail: andrea.rapp@tu-darmstadt.de

wie die Ergebnisse dieser Tätigkeiten – die Edition bzw. das Annotat – vornehmen zu können.

Die Edition (oder auch das „Edieren“) mit ihrer Doppelfunktion als Forschungs-Ergebnis und Grundlage weiterer Forschung dient zunächst der Dokumentation und Sicherung der Gegenstände (literatur-)wissenschaftlicher Betrachtung, darüber hinaus jedoch auch der Erschließung durch Kommentierung. Damit repräsentiert sie nicht eindeutig einen dieser zwei Pole, sondern lässt sich vielmehr als Kontinuum zwischen Sammlung bzw. Dokumentation und Interpretation beschreiben. Pointiert stellt sich die Frage, wo die Interpretation anfängt. Darüber hinaus ist Edieren, weil es Grundlagen schafft, immer auch methodischer und wissenschaftshistorischer Spiegel der Disziplin und ihres Verhältnisses zum Text sowie auch zu den Großkategorien Autor, Gattung, Epoche. In diesem Feld werden die elementaren Rollen von Produzierenden und Rezipierenden konzeptionalisiert, Autoritäten und Expertentum sowie das Lesen (an sich) oder Benutzen von Texten diskutiert bzw. teils radikal infrage gestellt und neu gedacht. Die Auseinandersetzung mit Digitalität führt demnach zu grundlegenden Reflexionen aller Kategorien und Praxeologien der Disziplin. Mit ihrem wissenschaftlichen Auftrag der Grundlagenschaffung ist die Editionsphilologie daher auch wichtiger Teil des Institutionalisierungsprozesses der Digital Humanities. Hier wiederholt sie die wissenschaftshistorische Rolle für die Philologien als Ganzes, obwohl Editionswissenschaft und Literaturwissenschaft durchaus nicht immer zusammengehen oder auch nur Notiz voneinander nehmen, was in verschiedenen Beiträgen der Sektion aufgegriffen wurde. Der Einzug des Digitalen hat auch das verwandte Thema der Digitalen Infrastrukturen aufs Tapet gebracht, das vielleicht nicht zuletzt so viel Aufmerksamkeit genießt, weil hier die Lücken und Desiderate immer noch schmerzlich groß sind, was digitale literaturwissenschaftliche Forschung nach wie vor in bestimmten Bereichen verhindert und in vielen anderen immer noch erschwert.

Die Annotation (oder auch „Annotieren“) als Praxis kann als eine zentrale Konstante von Text- und Wissenserschließung seit Beginn der Schriftlichkeit betrachtet werden. Annotieren zählt zu den sogenannten ‚Scholarly Primitives‘ (Unsworth)¹ und stellt eine elementare, aktive, mediale, ja körperliche Expression eines aktiven, aneignenden Leseprozesses dar. Insofern als Annotationen medial gefasst sind – gleich ob analog oder digital –, sind sie darüber hinaus dem analysierenden Zugriff verfügbar. In ihnen manifestiert sich vielleicht am deutlichsten die Hinwendung geisteswissenschaftlicher Forschung zum Prozesshaften, gleichsam in einer neuen Repräsentation des Hermeutischen Zirkels, der verstärkt das Ineinandergreifen von Vorwissen/Vorverständnis, Aushandlung, Dokumentation und Interpretation transparent macht.

Im Themenfeld Edition wurde nach wie vor das Verhältnis von analog und digital vielfältig adressiert, was verwundern mag, da das Geschäft des Edierens

¹ <http://www.people.virginia.edu/~jmu2m/Kings.5-00/primitives.html>.

in aller Regel seit langem digital betrieben wird. Insofern reibt sich die Editionsphilologie (und auch die Literaturwissenschaft?) bislang vor allem an einer Standortbestimmung medialer und funktionaler Aspekte, wie z. B. die Beiträge von Braun, Glauch und Kragl oder Nutt-Kofoth zeigen, wie sich die Zunft mit diesem Ablösungsprozess befasst. Diese Perspektiven werden entschieden erweitert in einer Diskussion, die sich um Leistungsfähigkeiten und Mehrwert digital-medialer Formate und digitaler Funktionalitäten dreht: skaliert die digitale Edition einfach nur gegenüber analogen Formaten oder liegt etwas fundamental Anderes vor? In diesem Zusammenhang lassen sich grundsätzlichen Fragen danach adressieren, was beispielsweise ein Kommentar oder ein Apparat leistet und wie man Apparate über ihre primäre Funktion der Dokumentation von Herausgeberentscheidungen hinaus für literaturwissenschaftliches Interpretieren fruchtbar machen kann. Ein Ansatz liegt beispielsweise darin, Kategorien für verschiedenen Kommentartypen zu bilden, um einen digitalen auswertenden Zugriff zu gewährleisten, was als Zeichen für die Abstraktionskraft, vielleicht auch den Modellierungssog digitaler Praktiken genommen werden kann.

Eine Diskussion darüber, ob der Begriff Digitale *Edition* noch adäquat ist oder ob wir nicht vielmehr über Wissensbasen, Wissensverbünde, Digitale Archive usw. sprechen sollten, wird nachdrücklich gefordert. Die interessante Frage nach den Unterschieden mediävistischer und neugermanistischer Überlieferung und damit verbundener Editionspraxis wurde gestreift, jedoch nicht vertiefend oder gar erschöpfend behandelt, würde aber reizvolle Möglichkeiten der weiterführenden Diskussion bieten.

Der Bereich von Apparat und Kommentar bahnt die Pfade zum Themenkomplex der Annotation über den Begriff der Explizierung. Eine wesentliche Eigenschaft von Annotationen ist ihre Diskretheit. Der Prozess des Annotierens lässt sich demnach als eine Kette von einzelstellenbezogenen, interpretativen Entscheidungen beschreiben. Damit hat das Annotieren Anteil an der modellierenden und operationalisierenden Komponente wie auch an der hermeneutisch-interpretierenden. Diese vor allem in den Beiträgen von Meister und Gius in den Blick genommene ‚Brückenfunktion‘ (oder auch Schnittstellenfunktion?) bedarf dringend weiterer Beschäftigung und Erforschung und kann eine zentrale Rolle bei der Standortbestimmung Digitaler Literaturwissenschaft spielen.

Die Beiträge belegen, dass Edition und Annotation im Kern der Veränderungen der Disziplin Literaturwissenschaft stehen, und dabei sowohl die Kontinuität literaturwissenschaftlichen Handelns als auch das immer wieder postulierte entscheidend Neue repräsentieren und sich daher jenseits der „Alltagspraxis“ als Reflexionsgegenstand anbieten. Die Beiträge zeigen darüber hinaus aber auch, welcher Erwartungsdruck auf Digitalität lastet und wie das konstruktiv zu kritischen Reflexionen führen kann, die zur Standortbestimmung sowohl der *Digitalen* Literaturwissenschaft als auch generell der *Literaturwissenschaft* beitragen.

Der Beitrag von Bauer, Viehhauser und Zirker (*Zwischenräume: Kommentierende Annotation und hermeneutische Bedeutungserschließung in digitalen Texten*) fragt nach der determinierenden Wirkung von Annotationen, insofern als sie eine interpretierende Abstraktionsebene darstellen. Es entsteht ein

Spannungsfeld zwischen der kanonisierenden Wirkung eines Kommentars und den Möglichkeiten digitaler Editionen, mehrere Textversionen parallel bieten zu können und damit ent-kanonisierend zu wirken. Vor diesem Hintergrund wird eine (auch im analogen vermisste) Theorie des Kommentars unter digitalen Bedingungen gefordert. Braun, Glauch und Kragl (Unterwegs zum Text ohne Herausgeber und ohne Leser. Eine medienpragmatische Standortbestimmung der digitalen Edition) nehmen das Verhältnis von Medium, Editor*in und Leser*in/Rezipient*in in den Blick und regen eine Begriffsschärfung für die digitale Edition an, um den fundamentalen Unterschied zwischen analogen und digitalen Erzeugnissen zu markieren. Im Digitalen wird der/die Editor*in transparent und der/die Rezipient*in gewinnt an Autorität, wird dadurch jedoch auch stärker gefordert. Brüning (Modellierung von Textgeschichte. Bedingungen digitaler Analyse und Schlussfolgerungen für die Editorik) zeigt Wege auf, wie der Apparat nicht mehr ‚Variantenfriedhof‘ bleibt, sondern – wenn er entsprechend aufbereitet ist für maschinelle Auswertungen – effizient interpretationsrelevante Sichten auf die Textgenese erlaubt. Brüning zeigt den Ansatz an neugermanistischen Beispielen auf, es scheint naheliegend, dass die Analyse älterer (handschriftlicher) Textüberlieferung besonders von diesen Überlegungen profitieren könnte und damit digitale Ansätze zur methodischen Brücke zwischen den Teildisziplinen werden. van Hulle (Notebook Editions: Creative and Selective Undoing of Source Texts) fügt den Überlegungen am Beispiel der Edition der Notizbücher Samuel Becketts weitere Facetten hinzu, indem das Verhältnis von Notizen und Entwürfen zur Kanonizität des Werks in den Blick genommen wird. Eine digitale ‚Archiv-Edition‘ und Kollationierungs-Tools, die von den Notizen und nicht vom Endprodukt Text ausgehen, erlauben evolutionäre Betrachtungen.

Meister (Annotation as Markup avant la lettre) weist nachdrücklich auf das Desiderat des Einsatzes digitaler Methoden in der Texthermeneutik hin, vor allem im Bereich ‚Interpretation‘. Eine Erklärung dafür könnte einerseits darin liegen, dass Formalisierbarkeit und Operationalisierbarkeit im Hinblick auf literaturwissenschaftliche Deutungsansätze als defizitär empfunden wird – zugleich Potenzial und Leistungsfähigkeit digitaler Verfahren falsch eingeschätzt bzw. mit Erwartungen überfrachtet werden. Gius (Digitale Hermeneutik. Computergestütztes *close reading* als literaturwissenschaftliches Forschungsparadigma?) knüpft an Überlegungen zur computergestützten literaturwissenschaftlichen Deutung an und führt Möglichkeiten im Bereich der Narratologie vor.

Nutt-Kofoth (Historisch-kritische Ausgabe digital. Eine Reformulierung der neugermanistischen Edition) schließt den Bogen mit grundsätzlichen Überlegungen zu einer Evolution der Editionsphilologie, die wie die Computerphilologie als Pfeiler einer literaturwissenschaftlichen Grundlagenforschung situiert ist. Er plädiert dafür, die digitale Edition als Knoten im Wissensverbund zu sehen, wobei die erschließenden Editionsteile als Vermittlungsbrücken zu den stärker interpretativen Anteilen der Wissensstätten fungieren, um Edition und Interpretation wieder stärker zusammenzuführen.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Zwischenräume. Kommentierende Annotation und hermeneutische Bedeutungserschließung in digitalen Texten

Matthias Bauer, Gabriel Viehhauser und Angelika Zirker

1 Einleitung

Annotationen gehören ohne Zweifel zu den charakteristischsten Techniken des Umgangs mit Texten im digitalen Zeitalter, und zwar auf so vielen Ebenen, dass sie vielleicht sogar überhaupt als gemeinsamer Nenner aller Bemühungen auf dem Gebiet der digitalen Textaufbereitung und -analyse angesehen werden können. Annotiert wird bei der computerlinguistischen Auswertung von grammatikalischen Textstrukturen und der Identifizierung von Features, bei der Aufbereitung von Texten für Editionen, bei der semantischen Textanalyse, beim Aufspüren von Entitäten im Text und auch bei der eher in der Soziologie verbreiteten Inhaltsanalyse.

Bereits John Unsworth hat das Annotieren zu der von ihm zusammengestellten Liste der „scholarly primitives“¹ in den Geisteswissenschaften gezählt, also zu jenen Techniken, die so grundlegend sind, dass sie über die Disziplinen hinweg als elementare Bausteine der geisteswissenschaftlichen Arbeit angesehen werden

¹John Unsworth, „Scholarly primitives. What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This“, in: *Humanities Computing. Formal Methods, Experimental Practice Symposium*, London 2000, <http://www.people.virginia.edu/~jmu2m/Kings.5-00/primitives.html> (letzter Aufruf 4.5.2018).

M. Bauer (✉) · A. Zirker
Eberhard Karls Universität Tübingen, Tübingen, Deutschland
E-Mail: m.bauer@uni-tuebingen.de

A. Zirker
E-Mail: angelika.zirker@uni-tuebingen.de

G. Viehhauser
Universität Stuttgart, Stuttgart, Deutschland
E-Mail: viehhauser@ilw.uni-stuttgart.de

können. *Annotating* erscheint bei ihm gleichberechtigt neben den allgemeinen Tätigkeiten des *Discovering*, *Comparing*, *Referring*, *Sampling*, *Illustrating* und *Representing*. Es stellt sich aber die Frage, ob dem Annotieren, zu dem Unsworth in seinem Überblick interessanterweise in der Folge deutlich weniger Worte verliert als zu anderen ‚primitives‘, nicht allein schon aufgrund seines geringeren Abstraktionsgrades eine besondere Rolle in dieser Aufstellung zugehört: Anders als etwa das Auffinden und Vergleichen von Textstellen ist Annotieren explizit an einen Schreibvorgang und eine entsprechende Tätigkeit gebunden. Annotationen sind nicht nur gedankliche Konzeptionen, sondern Texte, die an andere Texte herangetragen werden und sich dementsprechend auch materiell in Prätexte einschreiben.² Sie stehen damit in einem Mittelbereich zwischen Lesen und Schreiben, im wissenschaftlichen Kontext dann aber vor allem zwischen Primärtext und Analyse oder Interpretation. „Mesotexts“ hat Peter Boot daher Annotationen genannt,³ und in der Tat lässt sich die Annotation als die vielleicht bestimmende paratextuelle Form des digitalen Zeitalters bezeichnen.

Annotationen haben – als Text – dementsprechend immer auch eine mediale Komponente, was mit ein Grund dafür ist, dass sie als digitale Arbeitstechnik so bestimmend hervortreten, aber auch dafür, dass man sich mit ihnen unter dem Aspekt des Medienwandels näher beschäftigen sollte. In Arbeiten insbesondere zur semantischen Textanalyse ist immer wieder hervorgehoben worden, dass digitales Annotieren letztendlich die Fortsetzung der altehrwürdigen Kommentar- und Glossierungspraxis in analogen Büchern darstellt, die bis in die Antike zurückreicht und sogar bis zu der Verfestigung und Erstellung von neuen Texten geführt hat, die sich an den zu kommentierenden Text anlagern.⁴ Auch die literaturwissenschaftliche Arbeit selbst wird oftmals unter diesem Gesichtspunkt als Fortführung traditioneller, ans Buchmedium gebundener Arbeitstechniken wie dem Hervorheben von wichtigen Textstellen durch Unterstreichen oder dem Beifügen von Notizen am Seitenrand angesehen.⁵ Die Ubiquität der

²Zudem stehen Annotationen durch ihre Bindung an die Praxis des Schreibens logisch quer zu den anderen von Unsworth genannten Kategorien, die mitunter, wie etwa das *Discovering* und *Referring*, zum Bestandteil der annotierenden Tätigkeit werden können.

³Peter Boot, *Mesotext*, Amsterdam 2009, 13.

⁴Vgl. Andrea Rapp, „Manuelle und automatische Annotationen“, in: Fotis Jannidis/Hubertus Kohle/Malte Rehbein (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart 2017, 253–267, hier: 253; Thomas Bögel/Michael Gertz/Evelyn Gius u. a., „Collaborative Text Annotation Meets Machine Learning. heureCLÉA, a Digital Heuristic of Narrative“, in: *DHCommons* (2015), <https://dhcommons.org/journal/issue-1/collaborative-text-annotation-meets-machine-learning-heurecl%C3%A9-digital-heuristic> (letzter Aufruf 4.5.2018) sowie den Beitrag von Jan Christoph Meister in diesem Band.

⁵Trotz historischer und formaler Kontinuitäten besteht funktional freilich ein Unterschied zwischen ‚persönlichen‘ oder ‚privaten‘ und zur Veröffentlichung intendierten Annotationen, vgl. hierzu etwa Boot (Anm. 3), 42; Catherine C. Marshall/A. J. Bernheim Brush, „Exploring the Relationship between Personal and Public Annotations“, in: *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'04)* (2004), 349–357; Catherine C. Marshall, „The Future of Annotation in a Digital (Paper) World“, in: *GSLIS*, University of Illinois (1998), <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.20.8568&rep=rep1&context=1>

digitalen Annotation richtet also die Aufmerksamkeit auch auf diese im literaturwissenschaftlichen Betrieb oft schamhaft verschwiegene Tätigkeit des Lesens und Durcharbeitens, und zwar so stark, dass mitunter vernachlässigt wird, dass das Annotieren zumindest in der konventionellen Literaturwissenschaft ja nur eine Vorstufe der interpretatorischen Arbeit darstellt, die insbesondere durch die narrative Ausformulierung einer Interpretationserzählung ergänzungsbedürftig ist.⁶ Ähnlich wie bei der Wittgenstein'schen Leiter, die nach getaner Arbeit umgestoßen wird, werden die Vorarbeiten zumeist durch diese Interpretationserzählung aufgehoben, und nicht zuletzt deshalb mag heute noch Literaturwissenschaftlern der Gedanke, diese Annotationen als Forschungsdaten zu sammeln, ziemlich fremd vorkommen.

Jedoch ist das wahre Potenzial von Annotationen für digitale Methoden bekanntlich nur dann nutzbar, wenn diese nicht idiosynkratisch erfolgen, sondern Standards entsprechen, die die Annotationen dann auch intersubjektiv bzw. interoperational austauschbar und formalisierbar machen.⁷ Damit verschiebt sich die Aufmerksamkeit von der Interpretationsnarration auf die Annotationsarbeit, denn Standards und Formalisierung fördern die Abstraktion und nicht die übergreifende Erzählung. Jene digitalen Annotationsformen, die im Fokus der Diskussion der *Digital Humanities* stehen, sind dementsprechend meist punktuell und kurzgefasst, und nicht, wie manche der angerufenen Vorbilder aus der Kommentartradition, die als Annotationen bereits in die Narration übergreifen, ausschweifend und individuell.⁸

Durch diesen Zusammenschluss von Annotationstechniken und literaturwissenschaftlicher Arbeit, wie er durch die formalisierenden Methoden der *Digital Humanities* gefördert wird, kommt nun auch die Frage nach der determinierenden Wirkung von Annotationen und allgemein von Analysetechniken in den Blick.

[type=pdf](#) (letzter Aufruf 4.5.2018). Dieser funktionale Unterschied kann jedoch mitunter durch die Integration von ursprünglich personalen Annotationen in die Interpretationshypothese überwunden werden, weshalb auch für diese Art von Annotationen digitale Austauschplattformen konzipiert wurden, vgl. John Bradley, „Towards a Richer Sense of Digital Annotation. Moving Beyond a ‚Media‘ Orientation of the Annotation of Digital Objects“, in: *Digital Humanities Quarterly* 6 (2012), <http://www.digitalhumanities.org/dhq/vol/6/2/000121/000121.html> (letzter Aufruf 4.5.2018).

⁶Vgl. ähnlich auch Bögel/Gertz/Gius u. a. (Anm. 4): „it would be false to conflate the concepts of ‚annotation‘ and ‚interpretation““.

⁷Vgl. Rapp (Anm. 4), 257. Die Formalisierung fördert auch die Explizierung der zugrunde liegenden Konzepte, ebd. 256.

⁸Zum Problem der Zerstückelung (*Morselization*) und *Disaggregation* vgl. Willard McCarty, „A Network with a Thousand Entrances. Commentary in an Electronic Age?“, in: Roy K. Gibson/Christina Shuttleworth Kraus (Hg.), *The Classical Commentary. Histories, Practices, Theory*, Leiden 2002, 349–402, hier: 380, und die Beispiele unten in Abschn. 4.

In dem Moment, in dem Annotationen einen so hohen Stellenwert bei der literaturwissenschaftlichen Arbeit bekommen, dass sich die Modellierung von Texten in das Markup verlagert,⁹ stellen sich nicht nur Fragen der Standardisierung und Formalisierung einer früher mehr oder minder wild wuchernden Praxis, sondern auch sozusagen als Gegenbewegung die Frage, wie trotz all dieser Kategorisierung und Vereinheitlichung die potenzielle Bedeutungsfülle von Texten im Annotationsmodell bewahrt werden kann. Wenn die literaturwissenschaftliche Praxis wesentlich aus Annotation besteht, dann liegt der Gedanke nahe, nach Formen eines *Hermeneutic Markups* zu suchen, also nach einem Annotationskonzept, das formalisiert und trotzdem zugleich nicht-deterministisch ist und unterschiedliche Sichtweisen auf einen Text zulässt.¹⁰

Interessanterweise sind *Hermeneutic-Markup-Modelle* jedoch vor allem im Kontext von Annotationsansätzen erprobt worden, die sich auf formale oder narratologische Textanalysen beziehen. Diejenige Annotationsform aber, die der althergebrachten Tradition der Marginalglossierung vielleicht sogar am nächsten von allen Annotationsformen kommt und (im Sinne von Boots *Mesotext*) das vielleicht ausgeprägteste Bindeglied zwischen ‚objektiver‘ Primärtextanalyse und literaturwissenschaftlicher Interpretation darstellt, ist überraschenderweise im Kontext der *Digital Humanities* noch relativ wenig beachtet worden. Die Rede ist vom Kommentar, der wissenschaftlichen Editionen beigegeben wird und dort der Texterläuterung dient. Wir wollen im Folgenden zeigen, dass in der Diskussion um diese im Buchmedium etablierte Annotationsform bereits viele Fragestellungen vorweggenommen sind, die auch in der Reflexion über die literaturwissenschaftliche Geltung von *Digital-Humanities-Methoden* wiederkehren, was ein Nachdenken über diese spezifische Art von Markup – und überhaupt eine differenziertere Betrachtung von unterschiedlichen Annotationsformen – lohnend erscheinen lässt. Zudem wollen wir mit dem von Matthias Bauer und Angelika Zirker entwickelten System ‚TEASys‘ (Tübingen Explanatory Annotations System) ein Modell für eine Kommentarform diskutieren, das die Vorteile der Digitalisierung und Formalisierung von Annotationen nutzt, dabei aber gleichzeitig der Bedeutungsfülle von Texten gerecht werden will, sowie einige ausgewählte Probleme einer solchen Konzeption exemplarisch vorstellen.

⁹Dies hat etwa Patrick Sahle für digitale Editionen gezeigt: Ders., *Digitale Editionsformen*, Norderstedt 2013, bes. Bd. 2, 157–165.

¹⁰Wendell Piez, „Towards Hermeneutic Markup: an Architectural Outline“, in: *Digital Humanities 2010 Conference Abstracts*, London 2010, 202–205; Jan Christoph Meister, „Crowd Sourcing ‚True Meaning‘. A Collaborative Markup Approach to Textual Interpretation“, in: Marilyn Deegan/Willard McCarty (Hg.), *Collaborative Research in the Digital Humanities*, Farnham, UK, 2012, 105–122; Bögel/Gertz/Gius u. a. (Anm. 4); vgl. auch den Beitrag von Jan Christoph Meister in diesem Band.

2 Der Kommentar unter digitalen Bedingungen

Dass der Kommentar als Annotationsform in den Reflexionen über die *Digital Humanities* bislang noch selten Thema geworden ist, liegt wohl nicht zuletzt daran, dass diese Form der Texterläuterung auch in ihrer konventionellen Ausprägung als notorisch untertheorisiert gilt. Kommentieren entsteht in den meisten Fällen aus der Praxis und richtet sich nach den zumeist ganz unterschiedlichen Bedingungen der zu kommentierenden Texte. Reflexionen über das Kommentieren sind daher selten und verbleiben notwendigerweise auf einer sehr allgemeinen Ebene, schon allein aus dem Bewusstsein, dass nicht für jeden Textfall vorgesorgt werden kann.

Dennoch hat natürlich auch der Kommentar gewisse Zielsetzungen, die sich zum Teil implizit aus den Editionen erschließen lassen, zum Teil aber auch offen ausgesprochen werden. Als ein Grundprinzip des Kommentars kann wohl gelten, dass er das Verständnis eines Textes befördern soll, und zwar insbesondere jenes Verständnis, das durch die historische (oder kulturelle) Distanz zum zu kommentierenden Text verstellt ist. Der Text soll also vom heutigen Leser so gut verstanden werden können, wie ihn ein zeitgenössischer Leser hätte verstehen können. Worin jedoch genau diese Wissensdifferenz vom heutigen zum zeitgenössischen Leser besteht und wie ideal man sich letzteren vorzustellen hat, ist bereits eines der ersten Probleme, die der Kommentar aufwirft. Verfügte der zeitgenössische Leser tatsächlich über alle Wissenskontexte, die der Kommentar einer Ausgabe nachliefert, und wie weit darf man überhaupt bei der Erhellung der zu kommentierenden Texte gehen? Bei Siegfried Scheibe ist noch zu lesen, dass der Kommentar die Aufgabe habe, „zum *eindeutigen* Verständnis des Textes“ beizutragen.¹¹ Dieser Interpretationsoptimismus wird in späteren theoretischen Grundlegungen nicht mehr geteilt. Besonders deutlich wird dies etwa bei dem programmatischen Aufsatz von Gunther Martens mit dem Titel *Kommentar – Hilfestellung oder Bevormundung des Lesers?*,¹² der die selbst aufgeworfene Frage klar mit der zweiten Antwortalternative beantwortet und von dort aus die Kommentarpraxis grundsätzlich angreift: Kommentar ist Deutung des Textes, eine Umgangsform mit Texten also, gegen die grundsätzlich noch nichts einzuwenden wäre, allerdings erhalte diese Deutungsform in der kritischen Ausgabe durch die Beigabe zum Text kanonische Wirkung, die den Interpretationsansatz des Kommentars unzulässigerweise verabsolutiere:

Erhält nicht eine Interpretation automatisch einen anderen Status, sobald sie sich als Bestandteil einer kritischen Edition präsentiert? Deren Text soll dem Leser eine sichere Grundlage seiner Beschäftigung mit dem Werk bieten. Die Forderung nach höchster

¹¹ Siegfried Scheibe, „Zu einigen Grundprinzipien einer historisch-kritischen Ausgabe“, in: Gunter Martens/Hans Zeller (Hg.), *Texte und Varianten. Probleme ihrer Edition und Interpretation*, München 1971, 1–44, hier: 10 [Hervorhebung von uns].

¹² Gunter Martens, „Kommentar – Hilfestellung oder Bevormundung des Lesers?“, in: *editio* 7 (1993), 36–50.

Authentizität [...] weckt beim Leser eine Erwartung, die sich notwendig auch auf Kommentar und Erläuterung ausweitet. Anders als die Werkdeutung, die unter dem Namen des Interpreten in einer eigenständigen Veröffentlichung erscheint oder auch als Aufsatz in der Fachzeitschrift publiziert wird, erhält der Kommentar innerhalb einer Edition gleichsam die Dignität des kritisch hergestellten Textes: zuverlässig, abgesichert, authentisch.¹³

Es ist nicht zu übersehen, dass Martens' Argumentation medial gebunden erscheint und implizit von den Bedingungen des Druckmediums ausgeht: Die Dignität erhält der Kommentar durch die Integration „innerhalb einer Edition“, die eben ein anderes textliches Umfeld als eine Monographie oder eine Fachzeitschrift bietet. Es stellt sich daher die Frage, inwieweit Martens' Behauptung auch mit diesen medialen Voraussetzungen steht und fällt. Denn es gehört ja gerade zu den Eigenarten digitaler Editionen, dass diese zumeist nicht mehr den einen, gesicherten Text in den Mittelpunkt stellen, sondern in der Regel unterschiedliche Textversionen liefern, die ohne die räumlichen Beschränkungen des Drucks simultan präsentiert werden können. Damit haben digitale Editionen eine ent-kanonisierende und eine entgrenzende Wirkung,¹⁴ die so gesehen auf den Kommentar übergreifen sollte. „The electronic medium brings permanence [...] into question, and with it what we mean by authority, how it is established and maintained“, wie McCarty den Sachverhalt ausdrückt.¹⁵

Ist der Kommentar unter digitalen Bedingungen daher ähnlich neu zu denken wie die Edition? Digitale Editionen enthalten, wie Patrick Sahle herausgearbeitet hat, „fast zwangsläufig *mehr* als die traditionelle Edition“. ¹⁶ Dieses *Mehr* besteht insbesondere auch in einer Verschiebung von der Konzentration auf den Einzeltext (als linguistischen Code) hin zu den Kontexten, die „durch den Wegfall der Mengenbeschränkung und durch die grundsätzliche Netzstruktur der digitalen Publikation“ befördert werde.¹⁷ Fast scheint es so, als wäre also eine völlig entgrenzte Kommentierungspraxis (und damit auch die Auflösung der Grenze

¹³ Ebd., 46.

¹⁴ Die freilich auch wieder in neue Kanonisierungseffekte umschlagen kann, vgl. Sahle (2013), Bd. 2, 168–172.

¹⁵ McCarty (Anm. 8), 371. Die Auflösung der autoritativen Textgestalt war freilich schon in der vordigitalen Editorik als Tendenz erkennbar; dass damit entsprechende Konsequenzen für die Gestaltung des Kommentars verbunden sind, wurde bereits von Ulfert Ricklefs deutlich gemacht: Ders., „Zur Erkenntnisfunktion des literaturwissenschaftlichen Kommentars“, in: Wolfgang Frühwald/Herbert Kraft/Walter Müller-Seidel (Hg.), *Probleme der Kommentierung. Kolloquien der Deutschen Forschungsgemeinschaft Frankfurt am Main 12.–14. Oktober 1970 und 16.–18. März 1972*, Bonn-Bad Godesberg 1975, 33–74.

¹⁶ Sahle (Anm. 9), Bd. 2, 173.

¹⁷ Ebd.; Hans W. Gabler sieht dementsprechend gerade in einer diskursiven, über bloß positives Wissen hinausgehenden Kommentarpraxis das besondere Potenzial der digitalen Editorik, von der er sich eine Wiederannäherung von textkritischer und interpretativer Philologie erhofft. Vgl. ders., „Theorizing the Digital Scholarly Edition“, in: Ders., *Text Genetics in Literary Modernism and Other Essays*, Cambridge, UK, 2018, 121–142, hier: 139.

zwischen Laien und Experten) denkbar bzw. greifbar nahe, die noch dazu in der Lage ist, nicht nur *eine* Deutung des Textes nahezulegen, sondern ähnlich multiperspektivisch in den Dialog mit dem Text einzutreten wie die digitale Edition selbst. Eine solche Praxis ließe sich etwa mit dem Gedanken des *Semantic Web* verbinden, wo (zumindest auf den ersten Blick) alle Wissensbestände mit allen verknüpft sind,¹⁸ oder mit anderen Multiplizierungsmöglichkeiten der digitalen Welt wie etwa *Crowd-Sourcing-Verfahren*, welche die von Martens gefürchtete Bevormundung des Lesers mit den quasi-demokratischen Strukturen der berichtigten Schwarmintelligenz aushebeln könnte: Wenn jeder Nutzer seinen eigenen Blick auf den Text qua Annotation einbringen kann, dann ist automatisch davon auszugehen, dass die Pluralität der Deutungsmöglichkeiten, die ein Leser aus dem Text herausziehen kann, auch in dessen Kommentierung eingeht.

Doch ist zu fragen, ob eine solche Entgrenzung nicht derart an den Grundfesten des Kommentars rüttelt, dass sich die Form des Kommentars letztlich auflösen würde. Denn zu den weiteren Anforderungen, die immer wieder in theoretischen Grundlegungen genannt werden, gehört, dass der Kommentar nicht ausschweifend sein soll, sondern stattdessen immer möglichst textbezogen zu bleiben hat.¹⁹ Das Verhältnis von Kommentar und Text ist, so Karlheinz Stierle, eben keine beliebige intertextuelle Beziehung, sondern das „einer sachbezogenen und sich der Sache, dem kommentierten Text selbst, unterwerfenden Intertextualität“.²⁰ Gerade durch den Kommentar werden immer auch Textgrenzen definiert und Textprofile geschärft. Jan Assmann hat darauf hingewiesen, dass der Kommentar auch kanonisierende Wirkung auf den kommentierten Text selbst hat: Nur jene Texte, die es wert sind, werden überhaupt kommentiert, und die Glossierung wird vice versa zum Indikator für die herausgehobene Bedeutung des kommentierten Prätextes.²¹

Die Beschränkung auf den Textbezug ergibt sich also nicht nur aufgrund der materiellen Beschränkungen des zur Verfügung stehenden Platzes einer Buchedition (bzw. der damit verbundenen Unübersichtlichkeit der Verweismöglichkeiten mit steigender Komplexität), die medial durch die digitalen Vernetzungsmöglichkeiten aufgelöst werden könnte, sondern bis zu einem gewissen

¹⁸ Welche Möglichkeiten solche Verknüpfungen bergen, demonstriert der Beitrag von Stefan Dumont in diesem Band, der aber gerade wieder zeigt, dass Standardisierungen (und teilweise auch Komplexitätsreduktion) zu den unabdingbaren Voraussetzungen der digitalen Verbindung von Texten gehört. Zu Potenzialen und Auswirkungen des *Semantic Web* im Bibliothekskontext vgl. den Beitrag von Thomas Stäcker in diesem Band.

¹⁹ Vgl. zu diesen Forderungen etwa Christina Shuttleworth Kraus, „Introduction. Reading Commentaries/Commentaries as Reading“, in: Gibson/Shuttleworth (Anm. 8), 1–28, hier: 5.

²⁰ Karlheinz Stierle, „Werk und Intertextualität“, in: Ders./Rainer Warning (Hg.), *Das Gespräch*, München 1996, 139–150, hier: 149.

²¹ Jan Assmann, „Text und Kommentar. Einführung“, in: Ders./Burkhard Gladigow (Hg.), *Text und Kommentar. Archäologie der literarischen Kommunikation IV*, München 1995, 9–33, hier: 13. Vgl. hierzu auch Hans U. Gumbrecht, *The Powers of Philology. Dynamics of Textual Scholarship*, Urbana/Chicago 2003, 47.

Grad aus der traditionellen Prägung der Textsorte ‚Kommentar‘ selbst, die letztlich immer einen „reduktionistischen Charakter“ aufweist.²² Zugleich ist dem Kommentar aber immer auch ein Moment der Überschreitung inhärent, das sich aus der Zielsetzung der andauernden Aktualisierung der Verstehensbedingungen des zu kommentierenden Textes ergibt: Wenn es Aufgabe des Kommentars ist, die historische oder kulturelle Lücke zwischen dem Verständnishorizont, der für den Text veranschlagt werden kann, und dem aktueller Leser zu überbrücken, dann kann der Kommentar nie abgeschlossen sein, da immer neue Erklärungsschichten an den Text angelagert werden müssen, um den Text im Fortgang seiner Rezeptionsgeschichte verständlich zu machen.²³ Der Kommentar bewegt sich damit im Spannungsfeld von Offenheit und Geschlossenheit, zwischen Beschränkung und Erweiterung der Kontexte.²⁴ Im analogen Buchmedium wird diese Spannung durch den beschränkten Marginalraum, der für die Kommentierung zur Verfügung steht, austariert, der sich nun durch das digitale Medium potenziell öffnet.

Es stellt sich somit die Frage, ob den Kommentar im digitalen Zeitalter neu zu denken dann nicht auch hieße, die ‚vereindeutigende‘ oder auch die ‚Verständnis befördernde‘ Funktion des Kommentars in diesem Spannungsfeld neu zu denken. Angesprochen sind damit nicht zuletzt die Überlegungen, inwieweit die Methoden der *Digital Humanities* tatsächlich einen neuen Blick auf die Objekte der Geisteswissenschaften eröffnen oder lediglich performativere Verfahren zur Umsetzung des Althergebrachten mit sich bringen. Walter Morgenthaler hat letztere Möglichkeit, „Technik [...] als bloßes Mittel zur bequemeren Erreichung dessen, was man sowieso schon ohne sie wollte“, als „fantasielos“ bezeichnet und stattdessen die Potenziale eines „kreativen Zusammenspiels“ und des „Lustmoment[es], welches nur direkte Berührung gewährt“, beschworen, die von einem neu gedachten digitalen Kommentar ausgehen könnten.²⁵ Konkret schlägt Morgenthaler vor, neben einem „objektiven, zur Gültigkeit tendierenden, all-

²²Wolfram Groddeck, „„Und das Wort hab ich vergessen“. Intertextualität als Herausforderung und Grenzbestimmung philologischen Kommentierens, dargestellt an einem Gedicht von Heinrich Heine“, in: Gunter Martens (Hg.), *Kommentierungsverfahren und Kommentarformen. Hamburger Kolloquium der Arbeitsgemeinschaft für germanistische Edition 4. bis 7. März 1992, autor- und problembezogene Referate*, Tübingen 1993, 1–10, hier: 10.

²³Vgl. Gumbrecht (Anm. 21), 42–44; Hans G. Senger, „Der Kommentar als hermeneutisches Problem“, in: *editio* 7 (1993), 62–75, hier: 72–75.

²⁴Vgl. zu dieser Grunddifferenz der Kommentarfunktionen Roger Lüdeke, „Kommentar“, in: *Kompendium der Editionswissenschaften* (2002), http://www.edkomp.uni-muenchen.de/CD1/frame_edkomp_RL2.html (letzter Aufruf 4.5.2018).

²⁵Walter Morgenthaler, „Der produktionsorientierte Stellenkommentar in der Computer-Edition“, in: Gunter Martens (Hg.), *Kommentierungsverfahren und Kommentarformen. Hamburger Kolloquium der Arbeitsgemeinschaft für germanistische Edition 4. bis 7. März 1992, autor- und problembezogene Referate*, Tübingen 1993, 251–255, hier: 253. Die Argumentation zielt damit in eine ähnliche Richtung wie der *Algorithmic-Criticism-Ansatz* von Stephen Ramsay, der den Computer zur spielerischen (jedoch regelgeleiteten) Potenzierung von Deutungsmöglichkeiten nutzen will (vgl. ders., *Reading Machines. Towards an Algorithmic Criticism*, Urbana 2011), und die Überlegungen von McCarty (Anm. 8), 397: „The activity which computing greatly, newly enhances is endless, serious play“.

gemein leser-orientierten Kommentar“ auch einen „subjektiven, rein arbeitsbezogenen, internen Kommentar“ der Edition beizugeben.²⁶ Damit sind letztlich ähnliche Formen wie die zu Beginn angesprochenen Annotationsarbeitstechniken aufgerufen, die den Unterstreichungen und Randnotizen im konventionellen Buchmedium entsprechen. Zugleich stellen sich damit aber auch dieselben Probleme, denn es dürfte zu diskutieren sein, inwieweit das Teilen von ungeordneten und vorläufigen Notizen ohne systematisierende Interpretationsnarration einen literaturwissenschaftlichen Mehrwert ergibt. Auf der einen Seite kann durch den vorläufigen und fragmentarischen Charakter des ‚subjektiven‘ Kommentars der Problemcharakter eines Textes und die damit verbundene prinzipielle Unabgeschlossenheit literaturwissenschaftlicher Deutungsarbeit explizit in Erinnerung gerufen werden, auf der anderen Seite könnte die tatsächliche praktische Verwertbarkeit dieser Notizen insbesondere dann an der schiereren Fülle des Materials scheitern, wenn der ‚subjektive‘ Kommentar in dem *Crowd Sourcing* ähnlichen Verfahren gleich in einer Vielzahl von Forscherstimmen potenziert wird, die den Text mit ihren Kommentierungen versehen. Die Notwendigkeit zur Beschränkung ergibt sich nicht allein aus der materiellen Begrenzung des Kommentars, sondern auch aus der beschränkten Verarbeitungskapazität seiner Leser.²⁷

Hilfsmittel gegen eine solche Unübersichtlichkeit wäre die Standardisierung von Annotationen, die es zudem erlaubte, die Anmerkungen zu formalisierbaren Kategorien zu bündeln und so die Grundlage für deren rechnerische Auswertung zu schaffen;²⁸ die Schwierigkeit dabei liegt wiederum im Austarieren der Vielgestaltigkeit der Formen und der abstrahierten Modellierung eines Kategorienschemas.

Eine bloße Aufteilung des Kommentars in die beiden Stufen ‚subjektiv‘ und ‚objektiv‘, wie sie bei Morgenthaler vorgesehen ist, greift wohl zu kurz, da sie Gefahr läuft, den ‚objektiven‘ Kommentar mit dem Nimbus des Authentischen, Eindeutigen zu versehen, den Martens als anmaßend kritisiert hat, demgegenüber der bloß ‚subjektive‘ Kommentar ohnedies an Bedeutung verlieren dürfte. Zudem gehört die Frage, ob sich überhaupt eine Entscheidung zwischen ‚objektiven‘ und ‚subjektiven‘ oder auch zwischen statisch-unveränderlichen oder dynamischen Kommentarteilen treffen lässt, ebenfalls zu den Kernproblemen der Kommentartheorie. Zweifellos lassen sich positive Daten eindeutig festhalten; wo genau die Grenze zwischen ‚noch objektiv‘ und ‚schon subjektiv‘ anzusetzen ist, dürfte aber wohl strittig sein und von Fall zu Fall verschieden.

²⁶ Morgenthaler (Anm. 25), 252.

²⁷ Morgenthaler (ebd.) ist sich des Problems durchaus bewusst und fordert daher, dass „auch der ‚subjektive Kommentar‘ in kommunikationsfähiger Form festgehalten“ wird.

²⁸ Gefordert ist also auch im Digitalen ein Kommentieren als „Teil der geregelten Rede über den Text“, vgl. Klaus Kanzog, „Historizität und Aktualität. Semiotische Probleme des Erläuterns und Kommentierens“, in: *editio* 7 (1993), 76–84, hier: 76.

Gefragt ist also ein differenzierteres Kategorienschema, das unterschiedliche Kommentarschichten sortiert, ohne ‚Authentisches‘ von ‚Nicht-Authentischem‘ zu trennen, das offen genug ist, um Platz zu schaffen für die Anlagerung von immer neuen Deutungsperspektiven, dabei aber auch ausreichend Struktur vorgibt, mit der der nötige Überblick bewahrt werden kann. In dieser Hinsicht ließen sich standardisierte Annotationstaxonomien geradezu als eine neue Form der paratextuellen Begrenzung verstehen, in die die bündelnde Kraft des beschränkten und zugleich beschränkenden Marginalraums im neuen Medium transformiert wird. Die digitale Annotation eröffnet so eine neue Chance zur Bearbeitung eines alten Problems des Kommentars: nämlich den Ausgleich zu schaffen zwischen seiner Offenheit und Geschlossenheit, zwischen der Festlegung von Bedeutung und dem Aufzeigen von Bedeutungspotenzialen. Damit könnte sich der digitale Kommentar aber zugleich als paradigmatisch erweisen für eine digitale Literaturwissenschaft insgesamt, die von der Spannung zwischen deterministisch-formalisierenden und hermeneutischen Zugängen geprägt ist.

3 Definitheit/Diskretheit und Verankerung von Annotationen im Kontext von Auswerten und Verstehen

Bei dem Versuch, durch entsprechende Kategorien subjektive mit objektiver Kommentierung zu verbinden, stellt sich in besonderer Weise die Frage nach der Funktion des Annotierens. Subjektives Annotieren als sichtbare Spur eines interaktiven Leseprozesses dient in der Regel der persönlichen Erschließung eines Werkes und kann, wenn diese nicht Selbstzweck ist, im oben genannten Sinn Vorstufe eines weiteren individuellen Arbeits- und Verstehensprozesses sein. Objektive bzw. objektivierende Annotation hat eine andere kommunikative Funktion, denn es geht hier in der Regel darum, dass Experten anderen Lesern unter Rückgriff auf gesichertes Wissen beim Verstehen bzw. der Arbeit mit dem betreffenden Text helfen. Auch subjektive Annotierung kann eine externe kommunikative Funktion haben, etwa eine fangemeinschaftsstiftende,²⁹ aber natürlich auch eine erklärende, wenn z. B. in einem Seminar persönliche Annotationen ausgetauscht und miteinander verglichen werden.³⁰ Darüber hinaus können subjektive Annotationen ihrerseits zum Gegenstand wissenschaftlicher Untersuchungen werden, etwa hinsichtlich von Rezeptionsgeschichte, Leseprozessen und Leserverhalten. Uns geht es demgegenüber um erklärende Annotationen, die Verstehensprozesse gezielt befördern wollen, und ihre Veränderung durch digitale Texte. Dabei spielt eine wesentliche Rolle, dass die Digitalität dazu einlädt, statt der einzelnen Interpretation eines einzelnen

²⁹Vgl. <https://genius.com/>.

³⁰Vgl. dazu das MIT-Projekt: <http://www.annotationstudio.org> (letzter Aufruf 4.5.2018).

Annotators eine Vielfalt von Interpretationen bzw. Aushandlungs- und Diskussionsprozessen abzubilden, die Basis des Expertenwissens zu verbreitern sowie statt eines idealtypischen Lesers³¹ eine reale Leserschaft anzusprechen.³²

Für die erklärende bzw. kommentierende Annotation generell, aber insbesondere für die digitale Annotation ist die Frage der Verankerung und Diskretheit zentral. Digitale Annotationen, die entweder selbst als Markup in einen digitalen Text eingefügt sind bzw. mit diesem in Form von Datenbankeinträgen verbunden sind, müssen folglich sowohl eindeutig dem Text zuzuordnen als auch in ihrer Funktion eindeutig sein. Hier ergibt sich eine Spannung zur prinzipiellen Uneindeutigkeit von Verstehensprozessen, die sich in gedruckten Kommentaren tendenziell leichter kaschieren bzw. adressieren lässt.³³ Die Notwendigkeit einer klar abgegrenzten Verankerung im Text und einer klar definierten Funktion resultiert aus der originären Rolle digitaler Annotation als einer digitalen Auszeichnung bzw. Aufbereitung von Texten. Ein Beispiel sind die komplexen, aber klar systematischen TEI-Regeln. Systematische Definition und Diskretheit sind unerlässlich, wenn Annotationen dazu dienen, Texte auszuwerten. Dies gilt insbesondere dann, wenn im Sinne des *Distant Reading* größere Textkorpora im Hinblick auf bestimmte Informationen und Fragestellungen (also z. B. einzeltextunabhängige Strukturen und Entwicklungen) untersucht werden sollen. Von der *Auswertung* ist das *Verstehen* von Texten zu unterscheiden. Letzteres ist Angelegenheit bzw. Ziel des konventionellen erklärenden Annotierens bzw. des Kommentars, und hier sind die Phänomene, da es um hermeneutische Prozesse geht, sowohl auf textlicher als auch auf kategorialer Seite miteinander verflochten und oft nicht streng voneinander zu trennen.

Beide Aspekte, das Auswerten und das Verstehen, die Definitheit und Diskretheit einerseits und die Verflochtenheit andererseits, werden durch die digitale erklärende Annotation miteinander in Verbindung gebracht, und das Verhältnis wird einem Klärungsprozess unterzogen. Vereinfacht gesprochen, kann die Digitalisierung des Kommentars bzw. der erklärenden Annotation (im Sinne der Auswertbarkeit) dazu führen, dass deren meist implizite Voraussetzungen und Annahmen sichtbar werden. Und umgekehrt kann die Einbeziehung des Kommentars in die digitale Annotationspraxis deutlich machen, dass Auswertung vielfach auf Verstehen (und

³¹L. D. Goulden, „Approaches to the Contextual Annotation of Nineteenth-Century Historical Fiction. Constructing an Ideal Reader’s Response“, in: *English Studies in Canada* 27/1–2 (2001), 141–163.

³²Matthias Bauer/Angelika Zirker, „Explanatory Annotation of Literary Texts and the Reader. Seven Types of Problems“, in: Dies. (Hg.), *International Journal of Humanities and Arts Computing. Special Issue* 11/2 (2017), 212–232.

³³Ein konkretes Beispiel hierfür ist Booths Kommentar in seiner Ausgabe von Shakespeares Sonetten (Shakespeare 1977), der den Beginn des 116. Sonetts („Let me not ...“) als „May I never“ ohne Hinweis auf eine mögliche Ambiguität annotiert (William Shakespeare, *Shakespeare’s Sonnets*, hg. von Stephen Booth, New Haven 1977, 384); gleichzeitig aber wird die der worterklärenden Paraphrase widersprechende Interpretation gegeben: „Let also suggests powerful beseeching“.

damit auf Interpretation) beruht. Wie Gius und Jacke auf der Basis der Annotation narratologischer Kategorien von Zeitstrukturen mittels CATMA (Computer Aided Textual Markup and Annotation) gezeigt haben,³⁴ dient ein Kategoriensystem der Klärung von interpretatorischen Widersprüchen und kann z. B. helfen herauszuarbeiten, wann divergierende Interpretationen tatsächlich auf objektive Texteigenschaften, insbesondere Ambiguitäten, zurückzuführen sind, und wann sie andere Ursachen haben, z. B. eine unscharfe Begrifflichkeit. Umgekehrt führt die Einbeziehung des Kommentars in digitale Textpraktiken tendenziell zu einer Klärung der jeweils im Hinblick auf eine Auswertung zu berücksichtigenden Textbedeutung und hilft dadurch, bei quantitativen Untersuchungen später kaum je wieder zu eliminierende Fehlannotationen zu minimieren. Bei der Auswertung größerer, nicht annotierter Textmengen kann die systematische erklärende Annotation von Beispieldaten dazu führen, ein Bewusstsein für mögliche Probleme in der Datenauswertung zu schaffen, etwa im Spannungsfeld von Wortfrequenz und Semantik, d. h. von Wort und Sache.³⁵

Auch das Auswerten kann dem Ziel dienen, einen (einzelnen) Text besser zu verstehen. In dieser Hinsicht ergänzen sich die verschiedenen Funktionen der digitalen Annotation: Nicht nur können erklärende Annotationen zum korrekten Markup im Hinblick auf auszuwertende Textkorpora führen, es können auch Auswertungen von textauszeichnenden Annotationen im Gesamttext oder in größeren Korpora zur erklärenden Annotation von spezifischen Textstellen beitragen. So kann etwa einerseits für eine digitale stilistische Untersuchung von Shakespeare-Texten entscheidend sein, ob ein *Trigram* auch tatsächlich dasselbe bedeutet,³⁶ was in der Regel nur im Rahmen einer erklärenden Annotation dargelegt werden

³⁴ Evelyn Gius/Janina Jacke, „The Hermeneutic Profit of Annotation. On Preventing and Fostering Disagreement in Literary Analysis“, in: Angelika Zirker/Matthias Bauer (Hg.), *International Journal of Humanities and Arts Computing. Special Issue 11/2* (2017), 233–254.

³⁵ Vgl. z. B. die Aussage „Shakespeare’s women are about 50 % more likely to speak of love than men“ im Erläuterungsteil „What is WordHoard“ als Kommentierung einer lemmatisierten Suche nach „love“, <http://wordhoard.northwestern.edu/userman/whatiswordhoard.html> (letzter Aufruf 4.5.2018); sie wird (falsch) zitiert von Matthew L. Jockers, *Macroanalysis. Digital Methods and Literary History*, Urbana 2013, 23. Abgesehen davon, dass hier der Wortgebrauch von „love“ mit dem Sprechen über Liebe verwechselt wird, wäre die Aussagekraft der statistischen Daten mittels einer Reflexion des Inputs zu erhöhen. So kann z. B. die erklärende Annotation einer Stelle, an der eine Figur auf gängige Liebestheorien rekurriert, dabei helfen, Daten im Hinblick auf eine Reflexion über Liebe auszuwerten (indem z. B. einschlägige Wortfelder einbezogen werden).

³⁶ Ein Beispiel ist der Gebrauch des *Trigrams* „let me not“, das etwa in *Measure for Measure* 2.1.219 („I advise you let me not find you before me again“) etwas völlig anderes bedeutet (nämlich „you must not“) als in *Hamlet* 1.2.146 „Let me not think on’t“, wo es „I must not“ bedeutet, oder in *Julius Caesar* 1.2.32, „Let me not hinder, Cassius, your desires“, wo es „I will not“ bedeutet (zitiert nach William Shakespeare, *The Norton Shakespeare*, hg. von Stephen Greenblatt u. a., New York 2008). Eine Annotation des mehrdeutigen Gebrauchs von „Let me not“ am Anfang des 116. Sonetts macht auf dieses Bedeutungsspektrum aufmerksam. Bei anderen stilometrischen Verfahren zur Autoridentifikation, etwa dem Gebrauch der häufigsten Funktionswörter, liegt die Verbindung zur erklärenden (semantischen) Annotation ferner; auch hier wäre aber zu überlegen, ob nicht durch erklärende Annotation ein kombiniertes Verfahren

kann, und umgekehrt kann die Aufdeckung von identischen Wortfolgen bei der Auswertung zeitgenössischer Korpora wertvolle Informationen für die erklärende Annotation der betreffenden Textstelle bieten.

Für die erklärende digitale Annotation als besondere Form eines *Hermeneutic Markup* stellen sich also mindestens zwei Bedingungen: Zum einen muss sie, damit sie für die Auswertung nutzbar bleibt und umgekehrt Ergebnisse der Auswertung digitaler Korpora leichter aufnehmen und berücksichtigen kann, klar strukturiert sein. Zum anderen muss sie, wie in Abschn. 1 ausgeführt, offen genug sein, um der auch lokal zu erklärenden Bedeutungsfülle eines Textes gerecht zu werden. Daraus ergibt sich die Idee eines mittelweiten Kategoriensystems, das überdies den Vorteil besitzt, zu einer inhaltlichen Verbesserung der erklärenden Annotation beizutragen, da es im Sinne eines heuristischen Werkzeugs Annotatoren nötigt, über die Funktion ihrer Ausführungen zu reflektieren, ohne sie jedoch durch ein zu detailliert fixiertes Annotationsschema einzuengen. Hier ist der Vorschlag des Tübinger Lehr- und Forschungsprojekts TEASys zu nennen,³⁷ das alle Annotationen nach Kategorien (*Categories*) und Ebenen (*Levels*) gliedert.³⁸ Während insbesondere in analogen Editionsformen überwiegend implizite Annahmen bezüglich der Wissensvoraussetzung beim Leser getroffen werden,³⁹ sollen in TEASys solche Entscheidungen offengelegt und systematisiert werden. Jede Annotation zu einer Textstelle umfasst bis zu drei Ebenen der Komplexität und bis zu acht jeweils einzeln digital aufzurufende Kategorien, von denen fünf (*Linguistic, Formal, Intratextual, Intertextual, Contextual*) die Art der Information bezeichnen, die zur Erklärung herangezogen wird. Eine weitere Kategorie, *Interpretive*, setzt immer mindestens eine der informierenden Kategorien voraus, wodurch zu einer Klärung des in Kommentaren häufig intransparenten Verhältnisses von Information und Interpretation beigetragen werden soll. Die Kategorie *Textual* bildet die Brücke zu editionsphilologischen Aspekten der Kommentierung, wobei hier (da es sich um ein Kommentarsystem und nicht um ein Editionssystem handelt) nur diejenigen textphilologischen Aspekte

erleichtert wird, bei dem die Frequenz der Funktionswörter mit semantisch idiosynkratischem Wortgebrauch zusammen untersucht wird. Gerade bei Verfahren der *Authorship Attribution* im Werk Shakespeares besteht das Problem, dass lediglich Frequenz-Varianzen im Vordergrund stehen, jedoch keine Bedeutungsdifferenzierungen berücksichtigt werden; siehe dazu Matthias Bauer/Angelika Zirker, „Shakespeare and Stylometrics. Old and New“, in: Anne-Julia Zwierlein/Jochen Petzold/Katharina Boehm u. a. (Hg.), *Anglistentag 2017. Proceedings*, Trier 2018, 31–38.

³⁷ Matthias Bauer/Angelika Zirker, „Whipping Boys Explained. Literary Annotation and Digital Humanities“, in: Ray Siemens/Kenneth M. Price (Hg.), *Literary Studies in the Digital Age. An Evolving Anthology* (2015), <https://dlsanthology.commons.mla.org/whipping-boys-explained-literary-annotation-and-digital-humanities> (letzter Aufruf 4.5.2018); Bauer/Zirker (Anm. 32).

³⁸ Siehe dazu auch den *Living Style Guide*, der als Grundlage für das Annotieren erstellt wurde und einer kontinuierlichen Revision unterliegt: http://www.annotation.es.uni-tuebingen.de/?page_id=958 (letzter Aufruf 4.5.2018).

³⁹ Bauer/Zirker (Anm. 32).

Berücksichtigung finden, die für die Bedeutungserklärung relevant sind. Die Kategorie *Questions* berücksichtigt die Dynamik und kollaborative Natur der Annotation und trägt dazu bei, mögliche Interpretationsentscheidungen transparent zu machen, aber auch etwa Literaturrecherchen im Sinne der Nachhaltigkeit zu dokumentieren. Während TEASys vor allem für literarische Texte im engeren Sinn gedacht ist, lässt es sich für andere zu erläuternde Textsorten (etwa historio-graphische Texte, Essays) leicht adaptieren.⁴⁰

Die Praxis der Arbeit mit TEASys hat auf ein weiteres theoretisches Problem im Zusammenhang mit dem Verhältnis von Auswerten und Verstehen aufmerksam gemacht, das auf die grundlegende hermeneutische Frage nach dem Verhältnis von Einzelannotation und Textganzem (siehe Abschn. 3) hinführt: Während ein Markup des Textes immer klare und diskrete Anker braucht, um eine Auswertung zu ermöglichen, ist es bei Fragen des Verstehens nicht immer einfach, den zu annotierenden Textteil (Buchstabe, Wort, Satzteil, Gedichtzeile usw.) klar zu bestimmen und einzugrenzen. Ein Bewusstsein für das Problem scheint bei existierenden digitalen Annotationen und in der Literatur zum Kommentar bislang kaum vorhanden,⁴¹ es ist aber ubiquitär. So stellt sich z. B. die Frage, wo in einem gereimten Gedicht oder Epos Erklärungen zum Reimschema anzubringen sind: Ist der ganze Text (bzw. die Reimwörter) zu markieren (was unpraktisch und unsinnig wäre) oder nur der erste Reim oder nur ggf. bedeutungstragende Abweichungen?

Ein weiteres Problem der Diskretheit besteht darin, dass an den jeweiligen Textstellen/Ankern verschiedene erklärungsbedürftige Elemente anzutreffen sind bzw. verschiedene Arten von Erklärung an derselben Stelle nötig sind und diese häufig nicht voneinander unterschieden werden. So ist z. B. in Ian

⁴⁰Anschlussfähig wird es auch durch die Möglichkeit, das Kategoriensystem als Ontologie zu reformulieren. Dies wird allerdings hier noch nicht thematisiert.

⁴¹Soweit wir sehen, werden Fragen der konkreten Verankerung nicht diskutiert; es wird nur der Unterschied von Einzelstellenkommentar und Sammelanmerkungen bzw. separater Darstellung übergreifender Zusammenhänge reflektiert, z. B. bei Wolfgang Frühwald, „Formen und Inhalte des Kommentars wissenschaftlicher Textausgaben“, in: Ders./Herbert Kraft/Walter Müller-Seidel (Hg.), *Probleme der Kommentierung. Kolloquien der Deutschen Forschungsgemeinschaft Frankfurt am Main 12.–14. Oktober 1970 und 16.–18. März 1972*, Bonn-Bad Godesberg 1975, 13–32, hier: 30 f., und Jochen Schmidt, „Die Kommentierung von Studienausgaben. Aufgaben und Probleme“, in: Wolfgang Frühwald/Herbert Kraft/Walter Müller-Seidel (Hg.), *Probleme der Kommentierung. Kolloquien der Deutschen Forschungsgemeinschaft Frankfurt am Main 12.–14. Oktober 1970 und 16.–18. März 1972*, Bonn-Bad Godesberg 1975, 75–89, hier: 81 f., sowie Gerhard Spellerberg, „Der Kommentar einer ‚Studienausgabe‘. Ausgewählte Werke Lohensteins im Rahmen der ‚Bibliothek deutscher Klassiker‘“, in: Hans G. Senger (Hg.), *Philologie und Philosophie. Beiträge zur VII. Internationalen Fachtagung der Arbeitsgemeinschaft Philosophischer Editionen (12.–14. März 1997, München)*, Tübingen 1998, 140–153, hier: 142. Wir bedanken uns bei Miriam Lahrsow für den entsprechenden Literaturüberblick; sie hat in ihrer Tübinger Masterthese *Digital Explanatory Annotations for Literary Texts. Possibilities – Practices – Problems – Prospects* (2017) auf das Problem hingewiesen: https://bibliographie.uni-tuebingen.de/xmlui/bitstream/handle/10900/80731/Lahrsow_MA-Thesis.pdf?sequence=1&i&isAllowed=y (letzter Aufruf 4.5.2018).

McEwans Roman *Sweet Tooth* der für viele Leser erklärungsbedürftige Satz zu lesen: „Unlike its boisterous infant cousin, the Booker, the Austen didn't go in for banquets.“⁴² Hier stellt sich z. B. die Frage, ob „Booker“ und „Austen“ separat oder als gemeinsamer Anker zu annotieren sind und wo der Unterschied zwischen der Information über den realen Literaturpreis und der Information über die Tatsache, dass es sich bei *the Austen* um einen fiktiven Literaturpreis handelt, dokumentiert wird. Das Problem der abgegrenzten Verankerung wird durch digitale Annotationen virulent. In gewisser Hinsicht scheint z. B. bei dem eben genannten Satz das gedruckte Medium der auf diskrete Annotation angewiesenen digitalen Annotationspraxis überlegen, wie das Beispiel der *Amazon-x-ray-Annotation* in der Kindle-Ausgabe des Romans zeigt.⁴³ Die im Rahmen von TEASys für solche Fälle diskutierten Lösungen bestehen darin, Hyperlinks zwischen den jeweiligen Annotationen einzufügen oder neben den Annotationen für die Einzelwörter eine weitere Annotation für den ganzen Satz anzulegen. In jedem Fall hilft die Reflexion der möglichen Verankerung aber auch dabei, die Technik der Mischung von realweltlichen und fiktiven Elementen in McEwans fiktionaler Welt zu erkennen und zu erklären.

Ein weiteres Beispiel für das Problem, zwei benachbarte Wörter sowohl einzeln als auch gemeinsam bzw. in ihrer Relation zueinander annotieren zu müssen, ist Hamlets berühmte Formulierung „A little more than kin, and less than kind“, die im Hinblick auf die Relation von Teil und Ganzem unten in Abschn. 4 diskutiert wird. Die Bedeutungen von „kin“ und „kind“ beeinflussen sich gegenseitig, sodass es wenig sinnvoll ist, hier nur Einzelwörter anklickbar zu machen und mit Einzelannotationen zu versehen. Zugleich ist es aber natürlich im Hinblick auf die Auswertbarkeit annotierter Texte essenziell, bei „kind“ zu dokumentieren, dass hier die natürliche Verwandtschaftsbeziehung im Spiel ist.⁴⁴ Wieder bietet es sich im digitalen Medium an, sowohl ein Einzelwort-Markup vorzunehmen als auch eine Annotation für die ganze Äußerung zu erstellen, die überdies durch das keineswegs klare Verhältnis von „more“ und „less“ zusätzlichen Erhellungsbedarf bekommt.

Zwischen den genannten Verankerungsproblemen von Phänomenen, die einen ganzen Text betreffen (wie das Reimschema) und solchen, die sich auf zwei benachbarte Ausdrücke beziehen, sind noch jene zu nennen, die syntaktische und

⁴² Ian McEwan, *Sweet Tooth*, London 2012, 318; vgl. Angelika Zirker/Matthias Bauer, „Guest Editors' Introduction. Explanatory Annotation in the Context of the Digital Humanities“, in: Dies. (Hg.), *International Journal of Humanities and Arts Computing. Special Issue 11/2* (2017), 145–152, hier: 146.

⁴³ Da die x-ray-Annotationen nur in einer Wikipedia-Verlinkung von Einzelwörtern bestehen, wird der Leser hier bei „Austen“ völlig im Stich gelassen bzw. in die Irre geführt.

⁴⁴ Zur Sache Martina Bross, „„Equivocation will undo us“? Wordplay and Ambiguity in Hamlet's First and Second Line“, in: Angelika Zirker/Esme Winter-Froemel (Hg.), *Wordplay and Metalinguistic/Metadiscursive Reflection. Authors, Contexts, Techniques, and Meta-Reflection*, Berlin 2015, 25–46.

andere nicht unmittelbar benachbarte Phänomene und entsprechende Bedeutungsrelationen betreffen. Ein Beispiel hierfür ist die Annotation von Shakespeares 23. Sonett, dessen erste Zeile „As an unperfect actor on the stage“ lautet. Natürlich liegt es nahe, für den heutigen Leser „unperfect“ zu erklären (als Form von *imperfect*), aber neben der reinen Wortbedeutung stellt sich hier die Frage der Aussagebedeutung, denn was ist eigentlich mit „unperfect actor“ gemeint? Die zweite Zeile bietet dafür zugleich erklärenden und verdunkelnden Kontext, „Who with his fear is put beside his part“, der also aus der Rolle fällt oder seinen Text vergisst, aber aufgrund welcher Furcht? Es geht hier nicht darum, diese Fragen zu beantworten, sondern zu zeigen, wie sich sowohl theoretisch als auch praktisch Verankerungsprobleme stellen. Im vorliegenden Fall bekommen sie durch das einleitende „As“ noch eine besondere Schwierigkeit: Erst in der fünften Zeile, „So I, for fear of trust, forget to say“ wird klar, dass es sich hier eindeutig um eine Vergleichspartikel handelt und nicht um eine den Sprecher als Schauspieler identifizierende Konjunktion. Dies hat Auswirkungen auf die Bedeutung der ersten Zeile, denn es macht einen großen Unterschied, ob das lyrische Ich *als* Schauspieler spricht oder *wie* ein Schauspieler. Die Annotation der ersten Zeile muss also mindestens die fünfte Zeile einbeziehen, damit klar wird, dass es sich nicht um ein Gedicht über Schauspieler handelt, sondern um die rhetorische Figur der Praeteritio, mit der der Sprecher sagt, was er sagen will, indem er mittels eines Vergleiches sagt, dass er es nicht sagen kann. Solche Herausforderungen an ein systematisches Annotieren sind wiederum nur durch eine Kombination aus Hyperlinks zwischen Einzelannotationen, einer visuell darzustellenden Überlagerung von Annotationen und einer diskontinuierlichen Verankerung zu meistern; durch letztere wird es möglich, eine Annotation auf mehrere separate Wörter und Wortgruppen zu beziehen.

4 Das Verhältnis von lokaler zu globaler Erklärung und Annotation

Die lokale Verankerung wie auch die Diskretheit der Annotationen führt unweigerlich zu weiteren Problemen. Zwei solcher Probleme sollen im Folgenden exemplarisch diskutiert werden: (1) das Verhältnis von Teil und Ganzem eines Textes hinsichtlich des Verstehensprozesses, d. h. die Voraussetzung, die Bedeutung des Ganzen verstanden haben zu müssen, um relevante Annotationen für ein individuelles Item erstellen zu können; sowie (2) die kontextbezogene Identität und Nicht-Identität von Annotationen zu einem gleichlautenden Lemma, d. h. die Relation der Items und Annotationen zueinander. Sind also lokale Erklärungen eines Textes für die Gesamtbedeutung relevant bzw. werden sie durch die Bedeutung des Gesamttextes gedeckt? Verändern sich Bedeutungen im Laufe eines Textes, etwa indem lokale Denotationen oder Anspielungen aufgrund globaler Bedeutungen ausgehebelt werden? Letzteres ist insbesondere im Falle

von Mehrdeutigkeiten denkbar, die im Verlauf des Textes erkennbar oder aufgelöst werden.⁴⁵

Ein Beispiel hierfür ist der Titel von Ngaio Marshs 1956 erschienenem Kriminalroman *Scales of Justice*: Man denkt, insbesondere im Kontext der Gattung, direkt an die Waage der Gerechtigkeit, also an die Wiederherstellung von Gerechtigkeit im Verlauf der Handlung, die angestrebt wird.⁴⁶ Doch stellt sich am Ende des Romans heraus, dass die Metapher von der Autorin reliteralisiert wird und es die Schuppen eines Fisches sind (Engl. *Scales*), die zur Auflösung des Falls und damit zur Gerechtigkeit führen, also ganz wortwörtlich zu *Scales of Justice* werden.⁴⁷ Die Frage nach den Verankerungen von Annotationen wurde bereits diskutiert (siehe Abschn. 2); hier stellt sie sich bezüglich einer Annotation zum Titel, und zwar im Hinblick auf die Bedeutung des Titels für den Gesamttext: Weist man direkt am Anfang in einer Titelannotation auf die Ambiguität hin und verrät damit im Sinne eines *Spoilers* den Ausgang der Handlung, was die Lektüre potenziell hinfällig macht? Oder deutet man auf die strategische Ambiguität des Titels erst am Romanende hin? Ist eine solche Annotation überhaupt relevant? Die letztgenannte Frage deutet darauf hin, dass stets die Funktion einer Annotation in Rechnung gestellt werden muss: Wenn es nur um den Ausgleich der historischen Differenz des Verständnisses geht, ist die Erläuterung hinfällig. Handelt es sich aber um eine Spracherläuterung, etwa für Nichtmuttersprachler, dann ist sie vermutlich nötig und Voraussetzung für das Verstehen des Wortspiels. Bei englischen Muttersprachlern kann man mit einiger Sicherheit davon ausgehen, dass ihnen die Bedeutung von ‚Scales‘ bekannt ist, während dies bei Lernern der Fremdsprache vermutlich nicht der Fall ist.⁴⁸ Gleichmaßen kann eine Annotation aber über die Wortbedeutung hinaus auf die wortspielerische Eigenschaft des Titels hinweisen, die somit einen Baustein der ästhetischen Qualität des Werkes hervorhebt (und, in der Umkehr, wieder für das nicht-hermeneutische Markup eines Textes etwa durch

⁴⁵Dieser Annahme liegt die Voraussetzung zugrunde, dass Annotationen gerade dann wichtig sind, wenn es um die Gesamtbedeutung eines Textes geht. Es handelt sich um eine offene Frage, die jedoch bislang nicht systematisch betrachtet wurde; Ansätze z. B. bei Judith Flanders, „Detailism, Digital Texts, and the Problem of Pedantry“, in: *TEXT Technology* 2 (2005), 41–70. Eine entscheidende Rolle spielt dabei die jeweilige Art bzw. Vorstellung von „Gesamtbedeutung“ und damit die Art des Verhältnisses von Detail und Ganzem (z. B. additiv oder charakteristisches Detail vs. typisches Ganzes),

⁴⁶Ngaio Marsh, *Scales of Justice*, London 1956. Die Auflösung des Verbrechens und Wiederherstellung von Ordnung gehört zu den gattungskonstituierenden Merkmalen; vgl. z. B. J. A. Cuddon, *The Penguin Dictionary of Literary Terms and Literary History*, London ³1992, 229.

⁴⁷Siehe Esme Winter-Froemel/Angelika Zirker, „Ambiguity in Speaker-Hearer-Interaction. A Parameter-Based Model of Analysis“, in: Susanne Winkler (Hg.), *Ambiguity. Language and Communication*, Berlin 2015, 283–339, hier: 315.

⁴⁸In der Annotationspraxis wird hier die Schwierigkeit der Trennung verschiedener Funktionen und Ziele der Erläuterung deutlich, die in TEASys jedoch aufgrund der Auszeichnung von Kategorien aufgehoben wird.

POS-Auszeichnungen o. ä. relevant sein kann, z. B. wenn man den Text auf seine ästhetischen Merkmale hin untersucht).

Gleichzeitig wird hier deutlich, dass die Gattung bei der Art der Kommentierung eines Textes eine Rolle spielt: Beim Kriminalroman stellt sich die Frage nach Inhalt und Verankerung von Annotationen und ihren Inhalten offensichtlich in besonderer Weise, weil bestimmte Plot-Elemente und ihre Rätselhaftigkeit Teil des Vergnügens der Lektüre sind, sprachliches Verstehen allerdings ebenfalls Voraussetzung für dieses Vergnügen ist.

Ein weiteres Problem, das mit der Distinktheit von Annotationen verbunden ist, besteht in der Annotation identischer Textelemente. Der Kontext- und Funktionsbezug von erklärenden Annotationen kann dazu führen, dass ein mehrfach im Text auftauchendes identisches Textelement unterschiedlich annotiert werden muss, obwohl es die gleiche grammatische Bestimmung hat. Ein Beispiel dafür findet sich in William Shakespeares *A Midsummer Night's Dream*, in dem das Lexem ‚shadow‘ in unterschiedlichen Bedeutungen vorkommt. In der Kollokation „swift as a shadow“ (1.1.144) etwa bezieht es sich auf etwas, das nur flüchtig existiert,⁴⁹ in Pucks Anrede von Oberon als „king of shadows“⁵⁰ auf die Dunkelheit (siehe *OED*, I.1) und das Reich der Nacht, in dem die Feen aktiv sind, und während der Inszenierung des Theaterstücks im Dialog von Theseus und Hippolyta („The best in this kind are but shadows“, 5.1.208) wird es ambig: Es bezieht sich sowohl auf die Ähnlichkeitsbeziehungen (II.6.a.) sowie auf die damals gängige Bezeichnung des Schauspielers als „shadow“ (6.†b.), die von Puck in seinem Epilog wieder aufgenommen wird („If we shadows have offended“, Ep. 1). Keine der Verwendungen ist identisch, d. h. jedes *Item* muss einzeln bezüglich seiner Bedeutung im Kontext reflektiert und entsprechend annotiert werden.

Gleichzeitig kann es auch zur gegenseitigen Beeinflussung quasi-identischer bzw. zusammenhängender Elemente kommen. In der zweiten Szene von William Shakespeares *Hamlet* etwa lesen wir folgenden Dialog:

KING. [...] But now my Cosin *Hamlet*, and my sonne.

HAM. A little more then kin, and lesse then kind.

KING. How is it that the clowdes still hang on you.

HAM. Not so much my Lord, I am too much in the sonne.⁵¹

Die Wörter „kin“ und „kind“ sind wortspielerisch durch Paronomasie verbunden, deuten beide auf Verwandtschaftsverhältnisse hin und sind damit quasi-synonym.⁵² Doch haben beide auch noch weitere Denotationen, die durch den

⁴⁹Vgl. *Oxford English Dictionary* (*OED*), <https://www.oed.com/>: „shadow, n.“ II.4.c.

⁵⁰Shakespeare (Anm. 36), 3.2.348.

⁵¹William Shakespeare, *Hamlet* Q2, London 1604, <http://internetshakespeare.uvic.ca/Annex/Texts/Ham/Q2/default/> (letzter Aufruf 4.5.2018), 1.2.244–247.

⁵²Vgl. Bross (Anm. 44), 30.

Kontext der Tragödie aufgerufen werden⁵³ und in einer erklärenden Annotation berücksichtigt werden müssten (s. o., Abschn. 2). Hinzu kommt, dass es für alle zehn weiteren (nominalen oder adjektivischen) Verwendungen von „kind“ im Stück zu reflektieren gilt, welche Denotation und welche Konnotationen für sie gelten, sodass sie entsprechend annotiert werden können. Im Kontext digitaler Editionen kommt hier nun eine neue Möglichkeit ins Spiel, die den Nutzen digitaler Verfahren deutlich macht: Es ist möglich, Annotationen miteinander zu verbinden, indem man sie mit einer Art Meta-Annotation versieht, die man als *Note-Tagging* bezeichnen könnte:⁵⁴ Für das *Hamlet*-Beispiel wäre denkbar, dass man „kin“, „kind“ sowie Verwandtschaftsbeziehungen damit so auszeichnet, dass Längsschnitte durch den Text abgebildet werden können. Das *Note-Tagging* würde so auf Basis der Annotationen stattfinden, nicht auf der des Textes, und damit bereits die Semantik der Items einschließen. Dieses Verfahren hat dann wieder Auswirkungen auf die Bedeutungszusammenhänge von Teil und Ganzem.

4.1 Die Veränderung durch digitale Texte und Verfahren

Dies sind Probleme und Fragen, die seit jeher die Annotation literarischer Texte betreffen, die aber bislang nicht systematisch reflektiert worden sind.⁵⁵ Sie gelten auch für die Annotation von Texten im Rahmen digitaler Editionen; gleichzeitig bringen aber digitale Editionen andere Lösungen mit sich, und zwar insbesondere im Hinblick auf die literaturwissenschaftliche Annotation.⁵⁶

Bedenkt man das obige Beispiel aus *Hamlet*, so wird an dieser Stelle auch deutlich, wie schwierig die *Distant-Reading-Nutzung* von Annotationen ist, sobald die Semantik ins Spiel kommt und es nicht mehr um die oberflächliche Annäherung an größere Strukturen geht: Es muss grundsätzlich in jedem Einzelfall die Bedeutung eines Wortes oder einer Phrase überprüft werden.

⁵³ Siehe ebd., 30: „the structural ambiguity caused by the missing subject in Hamlet’s line makes several interpretations of the line possible which depend on the meaning of ‚kind‘ as adjective as well as on different aspects of the situational context presented in this scene. However, the parallel construction of Hamlet’s line together with the parallelism in Claudius’s preceding line also prompts us to consider ‚kind‘ as noun meaning either ‚[t]he family, ancestral race, or stock from which one springs‘ (*OED*, ‚kind‘ n. II.12.), or ‚[a] race, or a natural group of animals or plants having a common origin‘ (‚kind‘ n. II.10.a).“

⁵⁴ Hier geht es somit um die Herstellung von Verbindungen und Kohärenz durch die Auszeichnung aufgrund semantischer Felder bzw. gemeinsamer formaler Aspekte (etwa die Auszeichnung von Ironie).

⁵⁵ Martens (Anm. 12).

⁵⁶ Es geht hier in erster Linie um hermeneutisches Markup und erst sekundär um weitere Möglichkeiten der digitalen Annotation, etwa hinsichtlich korpuslinguistischer Verfahren der Aufbereitung (etwa POS), siehe dazu z. B. Noah Bubenhofer, *Einführung in die Korpuslinguistik. Praktische Grundlagen und Werkzeuge*, <http://www.bubenhofer.com/korpuslinguistik/> (letzter Aufruf 4.5.2018).

Diese Beobachtung führt damit letztlich auch wieder zur Frage nach der Verankerung von Annotationen hinsichtlich des Verhältnisses von lokaler zu globaler Bedeutung, denn auch die Veränderung von lokaler Bedeutung im Gesamtkontext ist über *Distant-Reading-Verfahren* (bislang) nicht nachvollziehbar. Die digitale erklärende Annotation kann somit auch dazu dienen, Probleme solcher Verfahren aufzuzeigen, etwa wenn ggf. Wörter und Kollokationen einfach nur gezählt werden, ohne deren lokale Bedeutungsunterschiede zu berücksichtigen (vgl. oben Abschn. 3 zu „let me not“).

Die digitale Einbettung von Annotationen eröffnet somit neue Möglichkeiten, etwa die Nutzung von Annotationen als Korpora.⁵⁷ So lassen sich z. B. alle Annotationen zu „kind“ in einem Text und bezogen auf andere Texte aufgrund der Speicherung von Annotationen in einer Datenbank finden und zueinander in Beziehung setzen. Gleiches gilt für Meta-Annotationen, wie sie bezüglich des Beispiels aus *Hamlet* angedeutet wurden und etwa für den Roman von Ngaiio Marsh greifen könnten: Markiert man dort etwa alle wortspielerischen Ausdrücke im Text, so kann dies helfen, z. B. die ästhetische Qualität dieses Textes zu beurteilen oder zu einer Gesamtinterpretation zu gelangen.⁵⁸ Die Möglichkeit, Annotationen nach Kategorien zu filtern (wie etwa in TEASys; s. o.), führt ebenfalls zu einem Erkenntnisgewinn, und zwar nicht nur hinsichtlich des heuristischen Charakters der Kategorien, sondern auch der Oberthemen, Motive, Ambiguitäten etc. In diesen Bereichen sind wesentliche Vorteile des Digitalen nachvollziehbar.

4.2 Fallbeispiel analoge vs. digitale Edition: Francis Beaumont & John Fletcher, *The Knight of the Burning Pestle* (1607/1613)

Im Folgenden sollen die bislang geschilderten Probleme bezüglich des Verhältnisses der lokalen zur globalen Erklärung bzw. Annotation eines Textes anhand eines Beispiels konkretisiert werden. Bei unserem Fallbeispiel handelt es sich um eine englische Komödie aus der frühen Neuzeit: *The Knight of the Burning Pestle* von Francis Beaumont und John Fletcher entstand vermutlich um 1607 und wurde 1613 erstmals gedruckt. Im 19. Jahrhundert entstanden erste wissenschaftliche Editionen des Textes; eine digitale Edition liegt bislang nicht vor, wird aber auf der Grundlage von TEASys derzeit erstellt.

⁵⁷ Siehe dazu auch Dirk Roorda, „Practical Linguistic Annotation. The Hebrew Bible.“, in: Angelika Zirker/Matthias Bauer (Hg.), *International Journal of Humanities and Arts Computing, Special Issue 11/2* (2017), 276–287.

⁵⁸ Dies trägt wiederum zur Plausibilisierung bei und kann als Strategie zur Identifikation von Interpretationen dienen, siehe Fotis Jannidis, „Polyvalenz – Konventionen – Autonomie“, in: Fotis Jannidis/Gerhard Lauer/Matías Martínez u. a., *Regeln der Bedeutung: Zur Theorie der Bedeutung literarischer Texte*, Berlin 2003, 305–328, hier 309 f.; siehe auch Gius/Jacke (Anm. 34).

Die Komödie besteht aus einer *Induction* und fünf Akten: Ein Bürger („Citizen“) besucht mit seiner Frau das Theater und ist mit dem Stück, das aufgeführt werden soll, nicht einverstanden, was er während des Prologs kundtut. Viel lieber als ein Stück über *The London Merchant* möchte er eine Romanze sehen, in der sein eigener Stand durch Heldenhaftigkeit hervortritt. Deshalb soll der eigene Lehrling Ralph zum Helden der Handlung eines Stückes werden, das nun parallel zur Handlung des *London Merchant* entsteht: *The Knight of the Burning Pestle*. Dessen Handlung wiederum ist klassisch der Komödie verhaftet: Der reiche Händler möchte seine Tochter Luce an Master Humphrey, einen alternden reichen Mann, verheiraten, Luce aber liebt einen anderen, mit dem sie schließlich flieht, jedoch gefunden wird – und bei einem Kampf im Wald treffen dann die Handlungsstränge um Ralph, den „Knight of the Burning Pestle“, und den „London Merchant“ aufeinander.

Das Stück ist, so viel deutet sich eventuell bereits in dieser sehr kurzen Zusammenfassung an, eine Parodie auf die Romanzenliteratur und dabei insofern metadramatisch, als es als *Play-Within-A-Play* konzipiert ist. Im Folgenden werden einige beispielhafte Annotationen vorgestellt, die sich auf die beiden oben eingeführten Kernprobleme beziehen, also (1) auf das Verhältnis von Teil und Ganzem eines Textes sowie (2) auf die kontextbezogene Identität und Nicht-Identität von Annotationen zu einem gleichlautenden Lemma.

Der Titel bildet, analog zum Eingangsbeispiel des Kriminalromans, ein hermeneutisches Problem im Sinne von (1): Bei *The Knight of the Burning Pestle* handelt es sich um eine intertextuelle Anspielung auf Rittergeschichten und ihre Parodien, deren Relevanz man jedoch nur bei Kenntnis (und Verstehen) des ganzen Stückes erkennt. Eine Anmerkung zur Intertextualität findet sich bereits in der Ausgabe von Murch aus dem Jahr 1908:

Attempts to find an origin for this title have resulted in mere conjecture. Weber, in speaking of the play's general resemblance to Don Quixote, says: „Indeed the very name of the play seems to be taken from the Knight of the Burning Shield, though no doubt our poets may have derived the appellation from some ancient romance, as Shakespeare probably did the epithet of the Knight of the Burning Lamp, which Falstaff bestows on Bardolph.“ Cf. 1 *Henry IV.* 3. 3. Dyce (l. XXXIV) says the „title was perhaps suggested by that of an earlier (and not extant) play, *The history of the Knight in the Burning Rock.*“ This play was produced at Court at Whitehall in 1578-9. Cf. Cunningham, *Extracts from the Accounts of the Revels at Court*, 1842, p. 142.⁵⁹

Hattaway, in der Einleitung zu der von ihm herausgegebenen Edition des Stückes (und nicht in einer Titelannotation!), ergänzt Murch dahingehend, dass er auf die Tradition der iberoromanischen Romanzenliteratur verweist, insbesondere *Amadis de Gaul*: „In this last there appears a Knight of the Burning Sword, who

⁵⁹ Herbert Murch (Hg.), *The Knight of the Burning Pestle*, New York 1908, 105; siehe zudem Alexander Dyce (Hg.), *The Knight of the Burning Pestle*, London 1843–1846.

was probably the original of the Knight of the Burning Pestle with its obvious phallic references.“⁶⁰

Eine digitale Annotation kann hier inhaltlich wenig hinzufügen, hat aber einen entscheidenden Vorteil, nämlich die Verlinkung zu weiteren Passagen im Text, in denen Intertextualität und Parodie eine Rolle spielen. So nimmt die Frau des Bürgers („Wife“) in der *Induction*, also während der Verhandlung dessen, was gespielt werden soll, den Titel dahingehend auf, als sie fordert: „Let him kill a lion with a pestle; let him kill a lion with a pestle“.⁶¹ Diese Forderung trägt wiederum zu ihrer Charakterisierung bei, sodass sich hier ein Netz miteinander verbundener Annotationen ergeben kann, die im gedruckten Buch nicht abzubilden sind (im Sinne des oben beschriebenen *Note-Tagging* auf Grundlage der Annotationen). Damit kann wiederum veranschaulicht werden, wie Textteile zur Gesamtbedeutung beitragen.⁶²

Ein Item, das in unterschiedlicher Bedeutung mehrfach im Text vorkommt (2), ist das Verb „wink“. Insgesamt wird es dreimal im gesprochenen Text und einmal flektiert in einer Bühnenanweisung erwähnt. Das erste Vorkommen findet sich im zweiten Akt. Hier unterhalten sich Master Humphrey und sein Schwiegervater in spe (der reiche Händler – „the London Merchant“): Luce, die Humphrey heiraten soll, möchte mit ihrem eigentlichen Geliebten fliehen, macht sich aber Humphreys Naivität zu Nutze, indem sie behauptet, sie würde der Heirat mit ihm nur zustimmen, wenn er sie aus dem väterlichen Haus entführt. Dies bespricht nun Humphrey mit Luces Vater, um dessen Einwilligung zu erhalten:

Humphrey. Nay, she must whirl
From hence – and you must wink;⁶³

Die heute geläufigste Bedeutung von „wink“ ist ‚blinzeln‘,⁶⁴ doch aus dem Kontext sieht man, dass sie hier nicht zutreffen kann. Eine weitere Bedeutung, die heute obsolet ist, bis ins frühe 19. Jahrhundert aber die geläufigste war, ist: die Augen schließen.⁶⁵ Zitner annotiert entsprechend: „wink] pretend not to see“, wobei es sich also um eine metaphorische Deutung von „wink“ handelt.⁶⁶

⁶⁰Michael Hattaway (Hg.), *The Knight of the Burning Pestle*, London 1969/1986, xv.

⁶¹Ebd., „Induction“, 42 f.

⁶²Des Weiteren ermöglicht die digitale Annotation nach TEASys eine spezifische Kategorisierung: Für Leser des Textes, die Nicht-Muttersprachler sind, ist vermutlich eine Sprachklärung von „burning pestle“ mit all seinen Konnotationen hilfreich, die bei Hattaway (s. o.) nur angedeutet werden.

⁶³Sheldon P. Zitner (Hg.), *The Knight of the Burning Pestle*, Manchester 2004, 5 f.

⁶⁴Siehe *OED*, „wink“ v.¹ 2.

⁶⁵*OED*, „wink“ v.¹ †1.a.

⁶⁶Zitner (Anm. 63), 82. Neben Zitners Ausgabe ist Hattaway (1986) der einzige, die ‚wink‘ an dieser Stelle erklärt: „turn a blind eye“ (36, Anm. 6).

Das nächste Vorkommen von „wink“ findet sich im dritten Akt, wenn Ralph einen Riesen bekämpfen soll und dieser von einem Wirt beschrieben wird; der Zuschauer merkt sehr schnell, dass es sich bei dem Riesen Barbaroso um einen „barber“ („whose business it is to shave or trim the beards, and cut and dress the hair, of customers“⁶⁷) handelt:

And rushing forth, brings in the errant knight
 And sets him down in an enchanted chair.
 Then with an engine which he hath prepared
 With forty teeth, he claws his courtly crown;
 Next makes him wink, and underneath his chin
 He plants a brazen pece of mighty bord,
 [...] ⁶⁸

Zitner annotiert die Stelle wie folgt: „makes him wink] by sprinkling him with perfumed water; see ll. 377 ff. below“.⁶⁹ Die Bedeutung von „wink“ ist hier also wohl ‚blinzeln‘, wobei in der Annotation lediglich die Ursache dafür angegeben wird. Der Verweis auf „ll. 377 ff. below“ entpuppt sich hingegen als wenig hilfreich: Blättert man weiter, so wird zwar eine Rasur beschrieben, die von Barbaroso vorgenommen wurde, doch sucht man vergebens nach „wink“; das Verb findet sich in einer vorausgehenden Bühnenanweisung bei Zeile 368: „Enter Squire [TIM], *leading one winking, with a basin under his chin.*“ Zu dieser Bühnenanweisung gibt es wiederum eine Annotation von Zitner: „winking] see l. 381 ff. below.“ Und die Anmerkung zu 381 bezieht sich auf „water“ im Text, wozu Zitner schreibt: „probably perfumed“.⁷⁰ Die Vorteile einer digitalen Edition sind an dieser Stelle besonders einleuchtend: Einfache Verlinkungen zwischen den Einzeleinträgen ersparen das Blättern und ermöglichen einen schnelleren Überblick als die Buchausgabe, in der Verankerungen von Annotationen aufgrund der medialen Bedingungen nicht zu visualisieren sind.

Die letzte Erwähnung von „wink“ findet sich im *Epilog*, der von Wife, also der Ehefrau des Citizen, gesprochen wird: „*Wife.* [...] I refer it to your own discretions whether you will applaud him or no; for I will wink, and whilst you shall do what you will“.⁷¹ In keiner der existierenden Annotationen wird das Wort „wink“ im

⁶⁷ *OED*, „barber“ n. a.

⁶⁸ Zitner (Anm. 63), III. 244–249.

⁶⁹ Ebd. 113, Anm. 248. Bei Murch fehlt der Querverweis in seiner Annotation: „For the purpose of anointing his eyes with perfumed water. ‚Your eyes closed must be anointed therewith also.‘ – Stubbes, *Anat. of Abuses* 2. 50“ (Murch [Anm. 59], 196, Anm. 270). Der intertextuelle Verweis in seiner Annotation ist wenig hilfreich und eher enigmatisch, weil man sich als Leser nun die Frage stellt, worum es sich bei Stubbes‘ Text handelt. Es wäre zum einen die Relevanz des Verweises zu prüfen, zum anderen könnte man, im Falle der Relevanz, in der digitalen Annotation hier einen Link etwa zum Eintrag der British Library zu diesem Text einfügen (<https://www.bl.uk/collection-items/the-anatomy-of-abuses-by-philip-stubbes-1583>).

⁷⁰ Zitner (Anm. 63), 119 Anm. 381.

⁷¹ Zitner (Anm. 63), Ep. 8–10.

Kontext erklärt, obwohl es in der heute nicht mehr gebräuchlichen Bedeutung von „Augen schließen“ verwendet wird, und zwar in diesem Fall nicht metaphorisch wie im zweiten Akt. Diese Bedeutung ist heutigen Leserinnen und Lesern des Stücks nicht unbedingt geläufig, was eine Annotation erforderlich macht. Solche Fälle sind etwa für die Unterrichtspraxis von Belang,⁷² wo die Annotation bzw. die Annotierung von Texten als reflexives Verfahren und produktive Erschließungsmethode eingesetzt werden kann. Im Hinblick auf die Annotation gleichlautender Lemmata wird deutlich, dass eine einfache Kopie des früheren Eintrags nicht möglich ist, weil „wink“ zwar in einer ähnlichen Bedeutung, dabei aber literal verwendet wird. Während Luces Vater die Flucht ignorieren soll, bietet Wife den übrigen Zuschauern an, die Augen zu schließen, damit sie ihre akustische Reaktion nicht mitbekommt. Dass es sich bei dieser Quasi-Synästhesie um eine Anspielung auf Shakespeares *A Midsummer Night's Dream* handelt, wo ein ähnlich parodierter Charakter, nämlich Bottom, gleichermaßen mit seinen Augen hört,⁷³ könnte ebenfalls annotiert werden. Damit wird aber auch wieder die Bedeutung des Teils für das Ganze im Sinne des hermeneutischen Zirkels deutlich: Die Verwendung von „wink“ im *Epilog* trägt, ebenso wie die Forderung „Let him kill a lion with a pestle“ in der *Induction*, zur Charakterisierung von Wife bei. In einer digitalen Edition ist deshalb auch denkbar, dass automatisch Vorschläge und Verweise zu anderen für das Thema relevante Annotationen angeboten werden, in diesem Falle also ein Netzwerk von Annotationen, die für die Charakterisierung von „Wife“ relevant sind.

Weiterhin kann anhand dieses Beispiels der Verwendung von „wink“ nochmals hervorgehoben werden, wie Annotationen selbst zu einem Korpus werden können, das nicht nur bei der Erstellung weiterer Annotationen hilft, sondern auch für übergeordnete Fragenstellungen Daten liefert. Die Anlage und Speicherung der Annotation in einer Datenbank ermöglicht die Suche nach Items und Annotation. Bei „wink“ ist auf dieser Grundlage somit auch eine Verlinkung zu anderen Annotationen möglich, die auf in ähnlicher Weise verwendete Items verweisen. Der Nutzer der digitalen Edition wird im vorliegenden Fall auf Shakespeares 43. Sonett geführt, das mit der paradox anmutenden Zeile „When most I wink, then do mine eyes best see“ beginnt⁷⁴ und in der „wink“ in gleicher Weise wie von Citizen's „Wife“ verwendet wird. Digitale erklärende Annotation kann hier zu einem Instrument für sprachgeschichtliche Untersuchungen werden wie für

⁷²Vgl. dazu Angelika Zirker/Judith Glässer/Augustin Kelava u. a., „Kompetenzmodellierung im Fach Englisch. Literaturwissenschaft meets Psychometrie“, in: Toni Bernhart/Marcus Willand/Sandra Richter u. a. (Hg.), *Quantitative Ansätze in den Literatur- und Geisteswissenschaften. Systematische und historische Perspektiven*, Berlin 2018, 149–171.

⁷³„The eye of man hath not heard, the ear of man hath not seen“ (Shakespeare [Anm. 36], 4.1.204–05).

⁷⁴William Shakespeare, *Shakespeare's Sonnets*, hg. von Katherine Duncan Jones, London 2003, 197.

Fragen des Verhältnisses wörtlicher und metaphorischer Bedeutungen oder Konzeptualisierungen der Sinneswahrnehmung.

Ergänzung 2021

Im Folgenden möchten wir auf einige Publikationen verweisen, die nach Verfertigung unseres Beitrags zum Thema kommentierende und hermeneutische Annotation entstanden sind; unsere Liste erhebt dabei keinerlei Anspruch auf Vollständigkeit. Wir verweisen bei den Beiträgen jeweils darauf, ob sie in erster Linie kommentierend oder hermeneutisch relevant sind.

Heiniger, Anna Katharina/Reiter, Nils/Wiedmer, Nathalie u. a., „Kann man Ästhetik zählen? Systematische Annotation und quantitative Analyse von Erzählerbemerkungen in den *Isländersagas*“, in: Annette Gerok-Reiter/Matthias Bauer/Anna Pawlak u. a. (Hg.), *Andere Ästhetik. Grundlagen – Fragen – Perspektiven*. Berlin [in Vorbereitung].

(hermeneutisch): Der Beitrag stellt die systematische Annotation von Erzählerbemerkungen in den anonym überlieferten, mittelalterlichen *Isländersagas* (altnord. *Íslendingasögur*) vor und geht dabei vor allem der Frage nach, ob und inwieweit diese als Teil einer Literarisierungsstrategie wirksam werden und damit Aussagen über ein den *Isländersagas* möglicherweise inhärentes Konzept von Autorschaft ermöglichen. Die Annahme ist, dass sich auf Grundlage der Annotation aus der Menge der gesammelten Daten ein für jede Saga jeweils individuelles Profil in der Verwendung der Erzählerbemerkungen erkennen lässt.

Nantke, Julia/Schlupkothén, Frederik (Hg.), *Annotations in Scholarly Editions and Research. Functions, Differentiation, Systematization*, Berlin 2020.

Der Sammelband beleuchtet den Facettenreichtum des Annotationsbegriffs in der digitalen Forschung aus unterschiedlichen disziplinären Perspektiven und nach unterschiedlichen Parametern, die von theoretischen Überlegungen und Methodenreflexion bis hin zur konkreten Annotationspraxis reichen. Er stellt damit einmal mehr die Vielgestaltigkeit der Annotationen im digitalen Kontext aus, bemüht sich aber insbesondere auch darum, Berührungspunkte zwischen den einzelnen Konzeptionen und Zugängen aufzufindig zu machen.

In unserem Zusammenhang sind insbesondere folgende Artikel von Relevanz:

Franken, Lina/Koch, Gertraud/Zinsmeister, Heike, „Annotationen als Instrument der Strukturierung“, 89–108.

(hermeneutisch) Franken u. a. stellen die Eigenschaften zweier unterschiedlicher Annotationskonzeptionen heraus, die geradezu die äußersten Pole in der Spannweite möglicher Annotationstätigkeit darstellen: Zum einen ‚Produkt-orientierte‘ Annotationen, wie sie etwa in der Computerlinguistik zum Einsatz kommen, zum anderen ‚Prozess-orientierte‘ Annotationen, die in der Kulturanthropologie etwa im Rahmen der Grounded Theory eine lange Tradition haben. Die beiden Konzeptionen lassen sich den logischen Operationen der Deduktion („Produkt-orientiert“) und der Induktion sowie der Abduktion („Prozess-orientiert“) zuordnen und beruhen daher auf ganz unterschiedlichen Voraussetzungen; gleichwohl lassen sich in beiden Disziplinen Übergangsformen und Berührungspunkte finden.

Hinzmann, Maria, „Annotationsspiralen‘ und ‚hermeneutischer Zirkel‘?“, 131–156.

(hermeneutisch) Der Verortung von Annotationen im Spannungsfeld von Induktion und Deduktion widmet sich auch der Beitrag von Hinzmann, die mit Hilfe einer präzisen Ausdifferenzierung der Vorstellung des hermeneutischen Zirkels den Stellenwert von Annotationen in Prozessen der Re-Konstruktion und der Re-Modellierung herausarbeitet.

Horstmann, Jan, „Undogmatic Literary Annotation with CATMA“, 157–175.

(hermeneutisch) Jan Horstmann demonstriert anhand der im forText-Projekt entwickelten 6. Version des CATMA-Annotationstools, wie sich auch im Kontext digitaler Arbeitsprozesse Annotationen ‚undogmatisch‘ – und damit nicht an ein bestimmtes Kategorienschema gebunden – zum Einsatz bringen lassen. Damit wird ein niedrigschwelliger

Einstieg für traditionell hermeneutisch arbeitende Literaturwissenschaftler:innen geschaffen, zugleich aber auch die Möglichkeit einer schrittweisen Kombination von mehr oder weniger standardisierten (und damit besser oder schlechter maschinell prozessierbaren) Annotationsprozessen eröffnet.

Reiter, Nils/Willand, Marcus/Gius, Evelyn, „Die Erstellung von Annotationsrichtlinien als Community-Aufgabe für die Digitalen Geisteswissenschaften“, 325–350.

Berichtet über die Methode des Shared-Tasks zur interdisziplinären Ausarbeitung von Annotationsrichtlinien und deren maschineller Weiterverarbeitung, die die Expertisen von unterschiedlich stark geisteswissenschaftlich bzw. informatisch ausgerichteten Forscher:innen zusammenbringen kann.

Schlupkothén, Frederik/Schmidt, Karl-Heinrich, „‘Commentary’ and ‘Explanatory Note’ in Editorial Studies and Digital Publishing“, 351–371.

(kommentierend) Der Beitrag erarbeitet eine systematische Unterscheidung von kommentierenden und erklärenden Annotationen auf der Basis einer Differenzierung von unterschiedlichen Ebenen von Dokumentfunktionen.

Bauer, Matthias/Zirker, Angelika, „Understanding (Through) Annotations. Introductory Remarks“, in: *Connotations* 29 (2020), 34–47, <https://www.connotations.de/article/bauer-zirker-understanding-though-annotations-introductory-remarks/> (letzter Aufruf 26.10.2021).

(kommentierend) In diesem Beitrag beleuchten die Autoren sowohl die Eigenart als auch die hermeneutische Leistung von erklärenden Anmerkungen. Dabei geht es auch um Kriterien zur Beurteilung von Annotationen: Sie müssen von Relevanz für den Text bzw. die Textstelle sein, zu der sie hinzugefügt werden, und die annotierte Textstelle muss für die Gesamtbedeutung des Textes relevant sein, damit die Annotation sein Verständnis fördert.

Reiter, Nils/Pichler, Axel/Kuhn, Jonas (Hg.), *Reflektierte Algorithmische Textanalyse. Interdisziplinäre(s) Arbeiten in der CRETA-Werkstatt*, Berlin 2020.

Hieraus sind insbesondere folgende Beiträge in unserem Kontext relevant.

Pagel, Janis/Reiter, Nils/Rösiger, Ina u. a., „Annotation als flexibel einsetzbare Methode“, 125–141.

Es handelt sich bei diesem Beitrag um eine Meta-Reflexion auf die Möglichkeiten von Annotation als Mittel der Interpretation und Entwicklung von Theorien im Rahmen der *Digital Humanities*.

Barth, Florian, „Annotation narrativer Ebenen und narrative Akte“, 423–438.

(hermeneutisch) Auf der Grundlage von Genettes Begriff der „narrativen Ebene“ geht es in dem Beitrag darum, Ebenen wie auch „narrative Akte“ zunächst zu formalisieren und dann in eingebetteten Narrativen zur Annotation zu nutzen.

Ketschik, Nora/Krautter, Benjamin/Murr, Sandra u.a., „Zur Theorie von Erzählebenen und ihrer Annotation im digitalen Kontext“, 439–464.

(hermeneutisch): Auch in diesem Beitrag geht es um das Konzept der narrativen Ebene, hier im Kontext von SANTA (s. o.), und die Entwicklung von Annotationsrichtlinien. Die Einbeziehung von prototypischen wie auch abweichenden Fällen begünstigt die Erfassung des Problems sowohl hinsichtlich seiner Theorie wie auch der praktischen Anwendung.

Reiter, Nils/Willand, Marcus/Gius, Evelyn, „A Shared Task for the Digital Humanities Chap. 1. Introduction to Annotation, Narrative Levels and Shared Tasks“, in: *Cultural Analytics* 4/3 (2019), <https://culturalanalytics.org/article/11192-a-shared-task-for-the-digital-humanities-chapter-1-introduction-to-annotation-narrative-levels-and-shared-tasks> (letzter Aufruf 26.10.2021).

Willand, Marcus/Gius, Evelyn/Reiter, Nils, „A Shared Task for the Digital Humanities Chap. 3. Description of Submitted Guidelines and Final Evaluation Results“, in: *Cultural Analytics* 4/3 (2019), <https://culturalanalytics.org/article/11194> (letzter Aufruf 26.10.2021).

(hermeneutisch) Die Beiträge zeigen anhand verschiedener theoriegeleiteter Richtlinien zur Annotation von Erzählebenen die Wechselwirkung zwischen narratologischen Konzepten und Annotationspraxis.

Zirker, Angelika/Kirchhoff, Leonie/Lahrsow, Miriam, „Students as Digital Annotators of Shakespeare’s Sonnets“, in: *Digital English* (2018), <http://www.digitalenglish.com.au/2018/12/17/students-as-digital-annotators-of-shakespeares-sonnets/> (letzter Aufruf 26.10.2021). (kommentierend) Der Beitrag stellt ein Beispiel aus dem TEASys-Projekt vor, und zwar die Vorgehensweise von Studierenden beim Annotieren von Shakespeares *Sonnetten*, die aufgrund ihrer historischen Distanz insbesondere für Nicht-Muttersprachler:innen des Englischen als besonders ‚schwierige‘ Texte gelten.

Literatur⁷⁵

- Assmann, Jan, „Text und Kommentar. Einführung“, in: Jan Assmann/Burkhard Gladigow (Hg.), *Text und Kommentar. Archäologie der literarischen Kommunikation IV*, München 1995, 9–33.
- Barth, Florian, „Annotation narrative Ebenen und narrative Akte“, in: Reiter, Nils/Pichler, Axel/Kuhn, Jonas (Hg.), *Reflektierte Algorithmische Textanalyse. Interdisziplinäre(s) Arbeiten in der CRETA-Werkstatt*, Berlin 2020, 423–438.
- Bauer, Matthias/Zirker, Angelika, „Explanatory Annotation of Literary Texts and the Reader. Seven Types of Problems“, in: Dies. (Hg.), *International Journal of Humanities and Arts Computing. Special Issue 11/2* (2017), 212–232.
- Bauer, Matthias/Zirker, Angelika, „Shakespeare and Stylometrics. Old and New“, in: Anne-Julia Zwierlein/Jochen Petzold/Katharina Boehm u.a. (Hg.), *Anglistentag 2017. Proceedings*, Trier 2018, 31–38.
- Bauer, Matthias/Zirker, Angelika, „Understanding (Through) Annotations. Introductory Remarks“, in: Connotations 29 (2020), 34–47, <https://www.connotations.de/article/bauer-zirker-understanding-though-annotations-introductory-remarks/> (letzter Aufruf 26.10.2021).
- Bauer, Matthias/Zirker, Angelika, „Whipping Boys Explained. Literary Annotation and Digital Humanities“, in: Ray Siemens/Kenneth M. Price (Hg.), *Literary Studies in the Digital Age. An Evolving Anthology* (2015), <https://dlsanthology.commons.mla.org/whipping-boys-explained-literary-annotation-and-digital-humanities/>.
- Bögel, Thomas/Gertz, Michael/Gius, Evelyn u.a., „Collaborative Text Annotation Meets Machine Learning. heureCLÉA, a Digital Heuristic of Narrative“, in: *DHCommons* (2015), <https://dhcommons.org/journal/issue-1/collaborative-text-annotation-meets-machine-learning-heurecl%C3%A9a-digital-heuristic>.
- Boot, Peter, *Mesotext*, Amsterdam 2009.
- Bradley, John, „Towards a Richer Sense of Digital Annotation. Moving Beyond a ‘Media’ Orientation of the Annotation of Digital Objects“, in: *Digital Humanities Quarterly* 6 (2012), <http://www.digitalhumanities.org/dhq/vol/6/2/000121/000121.html>.
- Bross, Martina, „Equivocation will undo us‘? Wordplay and Ambiguity in Hamlet’s First and Second Line“, in: Angelika Zirker/Esme Winter-Froemel (Hg.), *Wordplay and Metalinguistic/Metadiscursive Reflection. Authors, Contexts, Techniques, and Meta-Reflection*, Berlin 2015, 25–46.
- Bubenhof, Noah, *Einführung in die Korpuslinguistik. Praktische Grundlagen und Werkzeuge* (2006–2017), <http://www.bubenhof.com/korpuslinguistik/>.
- Cuddon, J. A., *The Penguin Dictionary of Literary Terms and Literary History*, London ³1992.
- Dyce, Alexander (Hg.), *The Knight of the Burning Pestle*, London 1843–1846.
- Flanders, Judith, „Detailism, Digital Texts, and the Problem of Pedantry“, in: *TEXT Technology* 2 (2005), 41–70.

⁷⁵ Sämtliche digitalen Referenzen wurden letztmalig am 4.5.2018 eingesehen.

- Franken, Lina/Koch, Gertraud/Zinsmeister, Heike, „Annotationen als Instrument der Strukturierung“, in: Julia Nantke/Schlupkothén, Frederik (Hg.), *Annotations in Scholarly Editions and Research. Functions, Differentiation, Systematization*, Berlin 2020, 89–108.
- Frühwald, Wolfgang, „Formen und Inhalte des Kommentars wissenschaftlicher Textausgaben“, in: Ders./Herbert Kraft/Walter Müller-Seidel (Hg.), *Probleme der Kommentierung. Kolloquien der Deutschen Forschungsgemeinschaft Frankfurt am Main 12.–14. Oktober 1970 und 16.–18. März 1972*, Bonn-Bad Godesberg 1975, 13–32.
- Gabler, Hans W., „Theorizing the Digital Scholarly Edition“, in: Ders., *Text Genetics in Literary Modernism and Other Essays*, Cambridge, UK, 2018, 121–142.
- Gius, Evelyn/Jacke, Janina, „The Hermeneutic Profit of Annotation. On Preventing and Fostering Disagreement in Literary Analysis“, in: Angelika Zirker/Matthias Bauer (Hg.), *International Journal of Humanities and Arts Computing. Special Issue 11/2* (2017), 233–254.
- Goulden, L. D., „Approaches to the Contextual Annotation of Nineteenth-Century Historical Fiction. Constructing an Ideal Reader’s Response“, in: *English Studies in Canada 27/1–2* (2001), 141–163.
- Groddeck, Wolfram, „„Und das Wort hab ich vergessen“. Intertextualität als Herausforderung und Grenzbestimmung philologischen Kommentierens, dargestellt an einem Gedicht von Heinrich Heine“, in: Gunter Martens (Hg.), *Kommentierungsverfahren und Kommentarformen. Hamburger Kolloquium der Arbeitsgemeinschaft für germanistische Edition 4. bis 7. März 1992, autor- und problembezogene Referate*, Tübingen 1993, 1–10.
- Gumbrecht, Hans U., *The Powers of Philology. Dynamics of Textual Scholarship*, Urbana/Chicago 2003.
- Hattaway, Michael (Hg.), *The Knight of the Burning Pestle*, London 1969/1986.
- Heiniger, Anna Katharina/Reiter, Nils/Wiedmer, Nathalie u.a., „Kann man Ästhetik zählen? Systematische Annotation und quantitative Analyse von Erzählerbemerkungen in den *Isländersagas*“, in: Annette Gerok-Reiter/Matthias Bauer/Anna Pawlak u.a. (Hg.), *Andere Ästhetik. Grundlagen – Fragen – Perspektiven*. Berlin [in Vorbereitung].
- Hinzmann, Maria, „„Annotationsspiralen“ und ‚hermeneutischer Zirkel‘?“, in: Julia Nantke/Schlupkothén, Frederik (Hg.), *Annotations in Scholarly Editions and Research. Functions, Differentiation, Systematization*, Berlin 2020, 131–156.
- Horstmann, Jan, „Undogmatic Literary Annotation with CATMA“, in: Julia Nantke/Schlupkothén, Frederik (Hg.), *Annotations in Scholarly Editions and Research. Functions, Differentiation, Systematization*, Berlin 2020, 157–175.
- Hoy, Cyrus (Hg.), *The Knight of the Burning Pestle*, Cambridge 1966.
- Jannidis, Fotis, „Polyvalenz – Konventionen – Autonomie“, in: Fotis Jannidis/Gerhard Lauer/Matías Martínez u.a., *Regeln der Bedeutung. Zur Theorie der Bedeutung literarischer Texte*, Berlin 2003, 305–328.
- Jockers, Matthew L., *Macroanalysis. Digital Methods and Literary History*. Urbana 2013.
- Kanzog, Klaus, „Historizität und Aktualität. Semiotische Probleme des Erläuterns und Kommentierens“, in: *editio 7* (1993), 76–84.
- Ketschik, Nora/Krautter, Benjamin/Murr, Sandra u.a., „Zur Theorie von Erzählebenen und ihrer Annotation im digitalen Kontext“, in: Reiter, Nils/Pichler, Axel/Kuhn, Jonas (Hg.), *Reflektierte Algorithmische Textanalyse. Interdisziplinäre(s) Arbeiten in der CRETA-Werkstatt*, Berlin 2020, 439–464.
- Lüdeke, Roger, „Kommentar“, in: *Kompendium der Editionswissenschaften* (2002), http://www.edkomp.uni-muenchen.de/CD1/frame_edkomp_RL2.html (letzter Aufruf 4.5.2018).
- Marsh, Ngaio, *Scales of Justice*, London 1956.
- Marshall, Catherine C., „The Future of Annotation in a Digital (Paper) World“, in: *GSLIS*, University of Illinois (1998), <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.20.8568&rep=rep1&type=pdf>.
- Marshall, Catherine C./Bernheim Brush, A. J., „Exploring the Relationship between Personal and Public Annotations“, in: *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '04)* (2004), 349–357.

- Martens, Gunter, „Kommentar – Hilfestellung oder Bevormundung des Lesers?“, in: *editio 7* (1993), 36–50.
- McCarty, Willard, „A Network with a Thousand Entrances. Commentary in an Electronic Age?“, in: Roy K. Gibson/Christina Shuttleworth Kraus (Hg.), *The Classical Commentary. Histories, Practices, Theory*, Leiden 2002, 349–402.
- McEwan, Ian, *Sweet Tooth*, London 2012.
- Meister, Jan Christoph, „Crowd Sourcing ‚True Meaning‘. A Collaborative Markup Approach to Textual Interpretation“, in: Marilyn Deegan/Willard McCarty (Hg.), *Collaborative Research in the Digital Humanities*, Farnham, UK, 2012, 105–122.
- Morgenthaler, Walter, „Der produktionsorientierte Stellenkommentar in der Computer-Edition“, in: Gunter Martens (Hg.), *Kommentierungsverfahren und Kommentarformen. Hamburger Kolloquium der Arbeitsgemeinschaft für germanistische Edition 4. bis 7. März 1992, autor- und problembezogene Referate*, Tübingen 1993, 251–255.
- Murch, Herbert S. (Hg.), *The Knight of the Burning Pestle*, New York 1908.
- Nantke, Julia/Schlupkothén, Frederik (Hg.), *Annotations in Scholarly Editions and Research. Functions, Differentiation, Systematization*, Berlin 2020.
- Pagel, Janis/Reiter, Nils/Rösiger, Ina u.a., „Annotation als flexibel einsetzbare Methode“, in: Reiter, Nils/Pichler, Axel/Kuhn, Jonas (Hg.), *Reflektierte Algorithmische Textanalyse. Interdisziplinäre(s) Arbeiten in der CRETA-Werkstatt*, Berlin 2020, 125–141.
- Piez, Wendell, „Towards Hermeneutic Markup. An Architectural Outline“, in: *Digital Humanities 2010 Conference Abstracts*, London 2010, 202–205.
- Ramsay, Stephen, *Reading Machines. Towards an Algorithmic Criticism*, Urbana 2011.
- Rapp, Andrea, „Manuelle und automatische Annotationen“, in: Fotis Jannidis/Hubertus Kohle/Malte Rehbein (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart 2017, 253–267.
- Reiter, Nils/Willand, Marcus/Gius, Evelyn, „Die Erstellung von Annotationsrichtlinien als Community-Aufgabe für die Digitalen Geisteswissenschaften“, in: Julia Nantke/Schlupkothén, Frederik (Hg.), *Annotations in Scholarly Editions and Research. Functions, Differentiation, Systematization*, Berlin 2020, 325–350.
- Reiter, Nils/Pichler, Axel/Kuhn, Jonas (Hg.), *Reflektierte Algorithmische Textanalyse. Interdisziplinäre(s) Arbeiten in der CRETA-Werkstatt*, Berlin 2020.
- Reiter, Nils/Willand, Marcus/Gius, Evelyn, „A Shared Task for the Digital Humanities Chapter 1. Introduction to Annotation, Narrative Levels and Shared Tasks“, in: *Cultural Analytics 4/3* (2019), <https://culturalanalytics.org/article/11192-a-shared-task-for-the-digital-humanities-chapter-1-introduction-to-annotation-narrative-levels-and-shared-tasks> (letzter Aufruf 26.10.2021).
- Ricklefs, Ulfert, „Zur Erkenntnisfunktion des literaturwissenschaftlichen Kommentars“, in: Wolfgang Frühwald/Herbert Kraft/Walter Müller-Seidel (Hg.), *Probleme der Kommentierung. Kolloquien der Deutschen Forschungsgemeinschaft Frankfurt am Main 12.–14. Oktober 1970 und 16.–18. März 1972*, Bonn-Bad Godesberg 1975, 33–74.
- Roorda, Dirk, „Practical Linguistic Annotation. The Hebrew Bible“, in: Angelika Zirker/Matthias Bauer (Hg.), *International Journal of Humanities and Arts Computing. Special Issue 11/2* (2017), 276–287.
- Sahle, Patrick, *Digitale Editionsformen*, Norderstedt 2013.
- Scheibe, Siegfried, „Zu einigen Grundprinzipien einer historisch-kritischen Ausgabe“, in: Gunter Martens/Hans Zeller (Hg.), *Texte und Varianten. Probleme ihrer Edition und Interpretation*, München 1971, 1–44.
- Schlupkothén, Frederik/Schmidt, Karl-Heinrich, „‘Commentary’ and ‘Explanatory Note’ in Editorial Studies and Digital Publishing“, in: Julia Nantke/Schlupkothén, Frederik (Hg.), *Annotations in Scholarly Editions and Research. Functions, Differentiation, Systematization*, Berlin 2020, 351–371.
- Schmidt, Jochen, „Die Kommentierung von Studienausgaben. Aufgaben und Probleme“, in: Wolfgang Frühwald/Herbert Kraft/Walter Müller-Seidel (Hg.), *Probleme der Kommentierung. Kolloquien der Deutschen Forschungsgemeinschaft Frankfurt am Main 12.–14. Oktober 1970 und 16.–18. März 1972*, Bonn-Bad Godesberg 1975, 75–89.

- Senger, Hans G., „Der Kommentar als hermeneutisches Problem“, in: *editio* 7 (1993), 62–75.
- Shakespeare, William, *Hamlet Q2*, London 1604, <http://internetshakespeare.uvic.ca/Annex/Texts/Ham/Q2/default/>.
- Shakespeare, William, *Shakespeare's Sonnets*, hg. von Katherine Duncan Jones, London 2003.
- Shakespeare, William, *Shakespeare's Sonnets*, hg. von Stephen Booth, New Haven 1977.
- Shakespeare, William, *The Norton Shakespeare*, hg. von Stephen Greenblatt u.a., New York 2008.
- Shuttleworth Kraus, Christina, „Introduction. Reading Commentaries/Commentaries as Reading“, in: Roy Gibson/Christina Shuttleworth Kraus (Hg.), *The Classical Commentary. Histories, Practices, Theory*, Leiden 2002, 1–28.
- Spellerberg, Gerhard, „Der Kommentar einer ‚Studienausgabe‘. Ausgewählte Werke Lohensteins im Rahmen der ‚Bibliothek deutscher Klassiker‘“, in: Hans G. Senger (Hg.), *Philologie und Philosophie. Beiträge zur VII. Internationalen Fachtagung der Arbeitsgemeinschaft Philosphischer Editionen (12.–14. März 1997, München)*, Tübingen 1998, 140–153.
- Stierle, Karlheinz, „Werk und Intertextualität“, in: Karlheinz Stierle/Rainer Warning (Hg.), *Das Gespräch*, München 1996, 139–150.
- Unsworth, John, „Scholarly Primitives. What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This“, in: *Humanities Computing. Formal Methods, Experimental Practice Symposium*, London 2000, <http://www.people.virginia.edu/~jmu2m/Kings.5-00/primitives.html>.
- Willand, Marcus/Gius, Evelyn/Reiter, Nils, „A Shared Task for the Digital Humanities Chapter 3. Description of Submitted Guidelines and Final Evaluation Results“, in: *Cultural Analytics* 4/3 (2019), <https://culturalanalytics.org/article/11194> (letzter Aufruf 26.10.2021).
- Winter-Froemel, Esme/Zirker, Angelika, „Ambiguity in Speaker-Hearer-Interaction. A Parameter-Based Model of Analysis“, in: Susanne Winkler (Hg.), *Ambiguity. Language and Communication*, Berlin 2015, 283–339.
- Zirker, Angelika/Bauer, Matthias, „Guest Editors' Introduction. Explanatory Annotation in the Context of the Digital Humanities“, in: Dies. (Hg.), *International Journal of Humanities and Arts Computing. Special Issue* 11/2 (2017), 145–152.
- Zirker, Angelika/Glaesser, Judith/Kelava, Augustin u.a., „Kompetenzmodellierung im Fach Englisch. Literaturwissenschaft meets Psychometrie“, in: Toni Bernhart/Marcus Willand/Sandra Richter u.a. (Hg.), *Quantitative Ansätze in den Literatur- und Geisteswissenschaften. Systematische und historische Perspektiven*, Berlin 2018, 149–171.
- Zirker, Angelika/Kirchhoff, Leonie/Lahrsow, Miriam, „Students as Digital Annotators of Shakespeare's Sonnets“, in: *Digital English* (2018), <http://www.digitalenglish.com.au/2018/12/17/students-as-digital-annotators-of-shakespeares-sonnets/> (letzter Aufruf 26.10.2021).
- Zitner, Sheldon P. (Hg.), *The Knight of the Burning Pestle*, Manchester 2004.

Online-Ressourcen

Genius, <https://genius.com/>.

Annotation Studio, <http://www.annotationstudio.org>.

Living Styleguide for Annotating Literature, http://www.annotation.es.uni-tuebingen.de/?page_id=958.

Oxford English Dictionary (OED), <https://www.oed.com>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Unterwegs zum Text ohne Herausgeber und ohne Leser. Eine medienpragmatische und medientheoretische Standortbestimmung der digitalen Edition

Manuel Braun, Sonja Glauch und Florian Kragl

Nach einer längeren Phase der Retrodigitalisierung sind die Geisteswissenschaften dabei, zum nativ-elektronischen Edieren überzugehen und die Möglichkeiten des neuen digitalen Mediums derart zu nutzen, dass eine digitale Edition nicht mehr ohne (funktionale) Verluste in das Medium des Buchs gebracht werden kann.¹ Damit stellt die Digitalisierung der Edition, verstanden als „erschließende Wiedergabe historischer Dokumente“², ein Teilphänomen jenes Medienwandels dar, der gegenwärtig so gut wie alle Bereiche des menschlichen Lebens erfasst. Einerseits ist es charakteristisch für einen solchen Wechsel des Leitmediums, dass sich die von ihm bedingten Änderungen ohne expliziten theoretischen Vorlauf gleichsam wie automatisch einstellen bzw. einschleichen. Das zeigt sich auch in der Praxis vieler digitaler Editionsprojekte – unser eigenes³ ist in dieser Hinsicht keine Aus-

¹ Es geht also im Folgenden weder um digitalisierte Editionen noch um solche, die zwar im Netz veröffentlicht werden, sich konzeptionell aber noch am gedruckten Buch orientieren. Zu diesen Unterscheidungen vgl. Patrick Sahle, *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, 3 Bde., Norderstedt 2013, hier: Bd. 2, 58 f. Zur Definition der digitalen Edition vgl. auch ebd., 148–155.

² Ebd., 138, erläutert ebd., 138–148.

³ Manuel Braun/Sonja Glauch/Florian Kragl (Hg.), *Lyrik des deutschen Mittelalters. Digitale Edition*, <http://www.ldm-digital.de/> (letzter Aufruf 19.8.2017).

M. Braun (✉)
Universität Stuttgart, Stuttgart, Deutschland
E-Mail: manuel.braun@ilw.uni-stuttgart.de

S. Glauch · F. Kragl
Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Deutschland
E-Mail: sonja.glauch@fau.de

F. Kragl
E-Mail: florian.kragl@fau.de

nahme –, die zwar große Mühe darauf verwenden, ihrem Material angemessene technische Lösungen zu finden, diesen Horizont aber kaum einmal für grundsätzliche Überlegungen verlassen und fragen, was es denn eigentlich heißt, digital zu edieren.⁴ Entsprechend verschieden und punktuell reagieren sie auf die neue mediale Situation. Andererseits definieren sich neue Medien häufig im Rückgriff auf alte, sei es, dass sie diese nachahmen, sei es, dass sie sich von ihnen absetzen. Beide Verhaltensweisen lassen sich auch in der Editorik beobachten, die zwischen Anlehnung an und Ablehnung von philologischen Konzepten des Buchzeitalters schwankt.

Der folgende Beitrag zielt auf beide Punkte, indem er das editorische Tun in der digitalen Welt reflexiv begleiten möchte – entsprechende Bemühungen sind, sieht man von der (auch für uns) grundlegenden Studie von Patrick Sahle einmal ab,⁵ bislang vor allem von der angloamerikanischen Wissenschaft unternommen worden –⁶ und sich dabei auf ein Phasenmodell des Medienwandels insofern bezieht, als er zunächst die Neuerungen der Digital- durch eine Kontrastierung mit der Buchedition herausarbeitet. Dieser Vergleich bezieht sich erstens auf die Bedingungen, die der Edition von ihrem jeweiligen Medium vorgegeben werden (1.1 Mediale Vorgaben), zweitens auf das ihr zugrundeliegende editorische Konzept (1.2 Textkritik und Textherstellung) sowie drittens auf die Rolle, die die Edition ihren Rezipienten zuweist (1.3 Rolle des Rezipienten). Sodann wenden wir uns einem möglichen Zukunftsszenario zu (2 Zukunftsperspektiven der digitalen Edition), in dem sich die digitale Edition vom Modell des gedruckten Buchs befreit hat und gleichsam zu sich selbst gekommen ist.⁷ Dazu gehören ein

⁴ Den Mangel an grundsätzlicher Reflexion beklagt auch Sahle (Anm. 1), Bd. 2, 108 f.

⁵ Da sich unsere Beobachtungen und Überlegungen in vielen Punkten mit jenen von Sahle (Anm. 1), Bde. 1 und 2, decken, verzichten wir im Folgenden darauf, jede dieser Überschneidungen einzeln nachzuweisen.

⁶ Vgl. etwa Arianna Ciula/Francesco Stella (Hg.), *Digital Philology and Medieval Texts*, Ospedaletto 2007; Mats Dahlström, „Drowning by Versions“, in: *Human IT 4/4* (2000), <https://humanit.hb.se/article/view/174/187> (letzter Aufruf 19.8.2017); Matthew J. Driscoll/Elena Pierazzo (Hg.), *Digital Scholarly Editing. Theories and Practices*, Cambridge, UK, 2016, DOI 10.11647/OBP.0095; Richard J. Finneran (Hg.), *The Literary Text in the Digital Age*, Ann Arbor 1996; Hans W. Gabler, „Theorizing the Digital Scholarly Edition“, in: *Literature Compass* 7/2 (2010), 43–56; Kenneth M. Price, „Electronic Scholarly Editions“, in: Ray Siemens/Susan Schreibman (Hg.), *A Companion to Digital Literary Studies*, Malden, MA/Oxford/Carlton, Victoria 2007, 434–450; Peter L. Shillingsburg, *From Gutenberg to Google. Electronic Representations of Literary Texts*, Cambridge, UK, 2006; Edward Vanhoutte, „Defining Electronic Editions. A Historical and Functional Perspective“, in: Willard McCarty (Hg.), *Text and Genre in Reconstruction. Effects of Digitalization on Ideas, Behaviours, Products and Institutions*, Cambridge, UK, 2010, 119–144.

⁷ Dieses Phasenmodell des Medienwandels und seine Übertragung auf die Editionswissenschaft bei Sahle (Anm. 1), Bd. 2, 100–109, im Rückgriff auf Michael Giesecke, *Von den Mythen der Buchkultur zu den Visionen der Informationsgesellschaft. Trendforschungen zur kulturellen Medienökologie*, Frankfurt a. M. 2002, 271–274.

neues Profil des Editors, die Veränderung des Erstellungs- und Publikationsmodus digitaler Editionen, aber auch die Wirkung jener neuen digitalen Werkzeuge, die von digitalen Editionen provoziert werden oder diese selbst provozieren.

Zur Diskussion stellen wir die These, dass das digitale Medium völlig neue Formen der Erarbeitung und des Benutzens von Editionen definiert, die demnach als revolutionär⁸ bezeichnet werden können. Ob und inwieweit sie das sind, wollen wir ergebnisoffen prüfen, und entsprechend soll unsere Einschätzung des Veränderungsgeschehens nicht nur dessen Gewinne, sondern auch seine Verluste ausweisen. Beide, der Befund und die Bewertung, können nur tentativ sein, da wir uns mitten in einem fortschreitenden Medienwandel befinden, Beobachter- und Beobachtetenstandpunkt also zusammenzufallen drohen. In welche Richtung der eingeschlagene Weg weisen könnte, lässt sich aber vielleicht doch ein Stück weit extrapolieren, wenn man gegenwärtige Entwicklungen der Editionswissenschaft, der Digital Humanities und der Modi digitaler Textpräsentation betrachtet. Unsere Überlegungen richten sich dabei auf das Konzept und die Idee digitalen Edierens, nicht auf eine Bewertung konkreter Umsetzungsversuche. Letztere sind uns vielmehr Symptome der zu beschreibenden medialen Transformation.

Auch wenn unsere Beobachtungen und Überlegungen auf digitale Editionen überhaupt abzielen, führt unsere disziplinäre Herkunft aus der Literaturwissenschaft dazu, dass im Folgenden editorische Bemühungen um literarische Texte im Zentrum stehen.⁹ Wir konzentrieren uns dabei auf solche Projekte, denen entweder aufgrund ihrer technischen bzw. methodischen Konzeption oder aufgrund ihres Gegenstands eine gewisse Exemplarität zugeschrieben werden kann.¹⁰

⁸Vgl. Shillingsburg (Anm. 6), 4.

⁹Zu den Differenzen zwischen historischen und philologischen Editionen in wissenschaftsgeschichtlicher Perspektive Patrick Sahle, „Die disziplinierte Edition – eine (kleine) Wissenschaftsgeschichte“, in: Matthias Thumser/Janusz Tandeci (Hg.), *Editionswissenschaftliche Kolloquien 2005/2005. Methodik – Amtsbücher, Digitale Edition – Projekte*, Toruń 2008, 35–52. Dass die Forderung, Editionen offen für interdisziplinäre Nutzung zu gestalten, in der Praxis an der Unzufriedenheit mit Kompromissen oder an einem zu großen Aufwand scheitert, arbeitet Sahle (Anm. 1), Bd. 1, 245–247, heraus.

¹⁰Ausgewertet sind: Bohnenkamp, u. a. (Hg.), *Johann Wolfgang Goethe. Faust. Historisch-kritische Edition*, Beta-Version 2, Frankfurt a. M./Weimar/Würzburg 2016, <http://beta.f Faust-edition.net/> (letzter Aufruf 19.8.2017); Marilyn, u. a. (Hg.), *Jane Austen's Fiction Manuscripts*, <http://www.janeausten.ac.uk/index.html> (letzter Aufruf 19.8.2017); Danielle Girard/Yvan Leclerc, *Les Manuscrits de Madame Bovary. Edition intégrale sur le web*, <http://bovary.fr/> (letzter Aufruf 19.8.2017); Braun/Glauch/Kragl (Anm. 3); Universitätsbibliothek Heidelberg/Universität Heidelberg, SFB ‚Materiale Textkulturen‘ (Hg.), *Welscher Gast digital*, <http://digi.ub.uni-heidelberg.de/wgd/> (letzter Aufruf 19.8.2017). Vgl. auch das Verzeichnis bei Patrick Sahle, *A Catalog of Digital Scholarly Editions*. v. 3.0. Snapshot (2008 ff.), <http://www.digitale-edition.de/> (letzter Aufruf 3.4.2017).

1 Die digitale Edition im Vergleich zur Buchedition

1.1 Mediale Vorgaben

Das Medium des gedruckten Buches macht einer Edition zahlreiche Vorgaben hinsichtlich der Menge und Anordnung der Inhalte. Seine Aufnahmefähigkeit ist in mehrfacher Hinsicht beschränkt. Bücher eignen sich am besten zur Wiedergabe von Texten; auch Bilder enthält das Buch seit seinen Anfängen, doch ist deren Reproduktion immer noch aufwendig, gerade wenn sie farbig sein sollen. Die Kapazität des Buchs insgesamt, aber auch die der Einzelseite sind begrenzt, weil Bücher weder beliebig dick noch beliebig groß werden können, wenn sie benutzbar und bezahlbar bleiben sollen. Das Hochformat und die mit ihm verbundene Vertikalität begünstigen die Konvention einer hierarchischen Anordnung der auf einer Seite platzierten Textteile, sodass etwa der Kommentar – im wörtlichen wie im übertragenen Sinne – unter dem (oder in älterer Zeit häufig rund um den) Editionstext steht. Das Nacheinander der Zeilen und Seiten formt einen linearen und sequenziellen Text; mehr als eine zusätzliche Dimension (etwa bei synoptischem Nebeneinander mehrerer Fassungen) ist kaum darstellbar. Die Unveränderlichkeit des Gedruckten vermittelt – verbunden mit der Einheitlichkeit der Codierung und der Ästhetik des Layouts, wie sie sich im Laufe der Buch- und Editions-geschichte herausgebildet haben – den Eindruck der Stabilität und Gültigkeit des Textes, sofern nicht mit hohem typographischem Aufwand diesem Eindruck entgegengearbeitet wird.

Demgegenüber ist die Kapazität des digitalen Speichermediums quasi unbegrenzt, und da es im Grunde auch nichts mehr kostet, entfallen alle Beschränkungen des Umfangs. Nicht nur Texte und Bilder in beliebiger Zahl und Größe, sondern auch Audio- und Videoaufnahmen können Teil der Edition sein.¹¹ Entsprechend enthalten digitale Editionen häufig nicht mehr nur den vom Herausgeber erstellten Text, sondern auch die Quellen in Gestalt digitaler Reproduktionen. Diese zielen auf die visuelle semiotische Dimension des Texts (Typographie, Layout, Material etc.) und dienen der Authentifizierung der Edition. Da die digitale Edition bei der Textdarstellung nicht mehr an das Format der Buchseite gebunden ist, kann sie einen Text nach den ihr als angemessen erscheinenden Kriterien gliedern und abweichende Fassungen theoretisch in unbeschränkter Vielfalt nebeneinanderstellen. Nur der Bildschirm setzt mit seiner Größe und seinem Format hier noch Grenzen – und zwar jeder Bildschirm eigene und andere –,¹² doch sind diese nicht von derselben Unüberwindlichkeit wie jene

¹¹Zu den neuen Möglichkeiten, aber auch Herausforderungen digitaler Editionen siehe Price (Anm. 6), 435–439.

¹²Zur Abhängigkeit der Textgestalt vom jeweiligen Ausgabegerät siehe Elena Pierazzo, „Modelling Digital Scholarly Editing. From Plato to Heraclitus“, in: Driscoll/Pierazzo (Anm. 6), 41–58, hier: 50 f. Joel Fredell, „Digital Philology in Little Boxes. Mobile Devices and *The Book of Margery Kempe*“, in: *Florilegium* 32 (2015), 77–98, macht darauf aufmerksam, dass das Smartphone bzw. das Tablet als Standardausgabegeräte die Rückkehr zum Buchformat bedeute; hierauf habe sich die digitale Editionsphilologie einzustellen.

einer gedruckten Buchseite. Auch Kontextmaterial kann jetzt Teil der Edition werden. Der Inhalt der Edition nimmt also zu, und die Möglichkeit der Verlinkung gibt Editionen offene Ränder, da diese mit beliebig vielen anderen Dokumenten verknüpft werden können.¹³ Umgekehrt kann eine digitale Edition unmittelbar von anderen Webseiten inkorporiert werden.¹⁴ Durch Hyperlinks, Harvesting durch Dritte und sonstige Vernetzungstechniken verliert die Edition ihre klare Kontur, das herausgegebene Werk den Charakter des In-sich-Geschlossenen.

Eine digitale Edition kann, anders als die gedruckte, schon insofern nie ein fester Text sein, als sie bei jedem Aufruf stets neu aus digitalen Daten generiert wird. Entsprechend teilt sich auch die Frage der Fortdauer der Edition in die nach der Oberfläche und die nach den Daten. Angesichts solch fundamentaler Differenzen stellt sich die Frage, ob man wirklich weiter von ‚Edition‘ sprechen oder nicht besser Begriffe wie ‚Informationsdatenbank‘ wählen will. Dass Texte aus einem Datenset ausgelesen werden, ermöglicht es, Editionen dynamisch zu gestalten¹⁵ und mehrschichtig anzulegen. Die Texte sind dann in unterschiedlicher Gestalt verfügbar, etwa als Transkription oder als Editionstext, womit die digitale Edition eine weite Skala von der Überlieferungs- zur Benutzernähe abdeckt. Auch die Zusammenstellung des Textkorpus, der Einstieg in dieses sowie die Anordnung der Texte auf dem Bildschirm sind potenziell frei konfigurierbar. Wenn die Buchseite insinuiert, sie repräsentiere den Text einfach, wie er ist, dann wird der Text auf dem Bildschirm ‚performativ‘. Die lineare Ordnung des gedruckten Buches wird – mit größerem oder (häufig) kleinerem Erfolg –¹⁶ ersetzt durch Listen, Register, Suchfunktionen sowie Strategien der Visualisierung.

1.2 Textkritik und Textherstellung

Mit dem gedruckten Buch untrennbar verbunden sind bestimmte Konzepte von Autorschaft, Textualität und Werkhaftigkeit: Texte gelten als abstrakte ‚platonische Ideen‘, die ‚jenseits ihrer materiellen Gestalt lieg[en]‘¹⁷. Indem traditionelle textkritische Editionen, diesem Prinzip folgend, Alternativen möglichst tilgen und mitunter drastisch bessern, was fehlerhaft oder sinnlos

¹³Price (Anm. 6), 435.

¹⁴Patrick Sahle, „Digitale Editionstechniken“, in: Martin Gasteiner/Peter Haber (Hg.), *Digitale Arbeitstechniken für Geistes- und Kulturwissenschaften*, Wien/Köln/Weimar 2010, 231–249, hier: 233.

¹⁵Zur Dynamik als Prinzip der digitalen Edition R. G. Siemens, „What Two Crowns Shall They Be?“ ‚Lower‘ Criticism, ‚Higher‘ Criticism, and the Impact of the Electronic Scholarly Edition“, in: W. Speed Hill (Hg.), *New Ways of Looking at Old Texts. III. Papers of the Renaissance English Text Society, 1997–2001*, Tempe, Arizona 2004, 37–46, hier: 41 f.

¹⁶Vgl. Dahlström (Anm. 6), 7 f.

¹⁷Sahle (Anm. 1), Bd. 1, 290; vgl. auch Pierazzo (Anm. 12), 46–49.

erscheint,¹⁸ richten sie das Überlieferte auf eine Weise zu, die es dem Medium des gedruckten Buches anverwandelt. Texte erscheinen unter Titeln und Autorangaben, die sie adressierbar machen, sie individualisieren und ihnen den Status von ‚Werken‘ verleihen. Jede Edition ist erst einmal final, allerdings kann sie durch eine Neuauflage oder Neuedition ihre Geltung verlieren. Dazu kommt, dass Bucheditionen vornehmlich aus Texten bestehen, während visuelle Strukturen und bildliche Elemente für sie von untergeordneter Bedeutung sind und sein müssen. Die Texte wiederum sind standardisiert, und zwar sowohl hinsichtlich ihrer Codierung als auch hinsichtlich ihrer sprachlichen Gestalt.

Die ‚klassische‘ Philologie – d. h. die Erschließung der kanonischen Schriften des Christentums und der Antike seit dem Humanismus und der Renaissance sowie die der ‚nationalsprachlichen‘ Literaturen, vor allem der des Mittelalters, seit etwa 1800 –¹⁹ hat nicht nur ganz auf die medialen Möglichkeiten des Buches gesetzt, sondern auch jene Textidee, die diesem Leitmedium gemäß ist, mit einer latent überzeitlichen Geltung versehen.²⁰ Aus der Perspektive der traditionellen Philologie erscheinen also auch ‚alte‘ Texte als Manifestationen jenes Literatursystems, das an ein vergleichsweise ‚junges‘ Medium, nämlich das gedruckte Buch, gebunden ist. Erst diese Setzung macht Autoren und/oder Herausgeber zu jenen ‚mächtigen‘ Instanzen der Textverantwortung, als die sie eine vorgutenbergische Galaxie nicht kennen konnte.

Natürlich bedeutet dies nicht, dass man nicht immer schon erkannt hätte, dass die spätantike und mittelalterliche Überlieferung etwa der Bibel, der Kirchenväter oder der mittelalterlichen Dichtung weit absteht von dem, was man von einem gedruckten ‚Werk‘ erwartet. Konsequenz dieser Beobachtung war aber im ‚langen‘ 19. Jahrhundert nie, die historische Differenz als solche gelten zu lassen, sondern vielmehr, sie über editorische Eingriffe zu minimieren. Was sich dann als editionsphilologische Techniken etabliert hat – Kollation, Rezension, Stemmabildung, Emendation, Konjekturen –,²¹ dient auch dem Zweck als dem, eine Überlieferung, die zu den medialen Gegebenheiten und Gepflogenheiten des gedruckten Buches völlig quer steht, eben diesen Gegebenheiten und

¹⁸ Zu den juristischen bzw. ethischen Untertönen des Anspruchs, gereinigte, fehlerfreie Texte vorzulegen, vgl. Gabler (Anm. 6), 52.

¹⁹ Vgl. Thomas Bein, *Textkritik. Eine Einführung in Grundlagen germanistisch-medievalistischer Editions-wissenschaft. Lehrbuch mit Übungsteil*, Frankfurt a. M. u. a. ²⁰¹¹ [2008]; zur Neueren deutschen Literatur: Herbert Kraft, *Editionsphilologie* [1990], Frankfurt a. M. u. a. ²⁰⁰¹.

²⁰ Hierzu und zum Folgenden Dahlström (Anm. 6), 3 f. Das heißt freilich nicht, dass die historisch-kritische Edition nicht auch andere historische Bedingungen hätte; schon der Wechsel von Editions-methoden innerhalb des Buchzeitalters spricht hierfür. Folglich sind mediale und mit ihnen verbundene ökonomische Faktoren ein wichtiger, aber nicht der einzige Faktor für die Theorie und Praxis der Edition. Dass die obigen Ausführungen diese komplexe Gemengelage reduzieren, ist dem Erkenntnisziel des Beitrags geschuldet.

²¹ Klassisch geworden ist die Darstellung von Paul Maas, *Textkritik*, Leipzig ¹⁹⁶⁰ [1927].

Gepflogenheiten zu unterwerfen. Gewiss hat man diesem Vorhaben seit jeher den Deckmantel wissenschaftlicher Rekonstruktion umhängen wollen, und gewiss ist dieser Deckmantel nicht immer ganz unpassend, nämlich dann, wenn Gegenstand der Edition Texte sind, die schon im Moment ihrer Entstehung auf derselben Textidee beruhen, wie sie die Edition voraussetzt.²² Dennoch kann dies nicht darüber hinwegtäuschen, dass die traditionelle Editionsphilologie, indem sie ihre buchkulturell geprägten Begriffe von Autor, Werk und Literatur zur Grundlage ihrer historischen Betrachtung gemacht hat,²³ in erster Linie auch – und meist ohne dies zuzugeben – Texte erstellt hat, die den Leseraugen ihrer jeweiligen Zeit vertraut sein mussten.

Für Texte jüngerer Epochen, deren Überlieferung in der Regel auch autorisierte Fassungen beinhaltet, gilt all dies analog. Zwar scheiden dann rekonstruierende Verfahren aus, weil es ihrer nicht bedarf, wenn ein letztgültiger Text erhalten ist; die Art aber, wie historisch-kritische Ausgaben die Genese dieses Textes (meist) letzter Hand ausfalten, zeigt ihre Herkunft aus der traditionellen philologischen Editorik. Präsentiert werden üblicherweise nicht alle verfügbaren Fassungen, sondern eine gültige Fassung – über deren Wahl sich dann streiten lässt –²⁴ samt einem Apparat, der die Textgeschichte aufbereitet. Je vielfältiger und je reichhaltiger das Material ist, desto schwerer fällt es jedoch, nachvollziehbare Kriterien dafür anzugeben, welche Varianten die ‚wichtigen‘ und also im Apparat wiederzugebenden sind. Gerade hier zeigt sich die Nähe zur Textkritik der Mediävistik und der Altphilologie, deren Apparate bei zu reicher Überlieferung ebenfalls überzulaufen drohen.²⁵ Solche praktischen Probleme sind Ausdruck des grundsätzlichen Dilemmas historisch-kritischer Editionen, die „zwischen Historizität

²² Belege dafür, dass Werktreue und Textstabilität nicht nur Konzepte der Neuzeit sind, wären etwa die Pflege antiker Literatur zur Blütezeit der Karolinger, ohne die diese nicht oder sehr viel schlechter erhalten wäre – man denke etwa an Alkuins textkritische Initiativen; siehe zusammenfassend Doris Haberl, „Die Hofbibliothek Karls des Großen als Kristallisationspunkt der Karolingischen Renaissance. Geschichte, Umfeld, Wirkungen“, in: *Perspektive Bibliothek* 3.1 (2014), 111–139 –, oder die (natürlich von der Heiligkeit des Gegenstandes stark aufgeladene, also nicht primär textphilologische) Legendenbildung um die *Septuaginta*, bekanntlich gründend im sogenannten *Aristeasbrief*: Aristeas, *Der König und die Bibel*. Griechisch/Deutsch, übersetzt und hg. von Kai Brodersen, Stuttgart 2008 [ca. 2. Hälfte des 2. Jahrhunderts v. Chr.].

²³ So erklärt sich etwa, dass – im neugermanistischen Bereich – die Aktualisierung von Zeichensetzung und Orthographie seit jeher als problematisch empfunden worden ist. Vgl. Kraft (Anm. 19), 72–92.

²⁴ Ein Beispiel für solche Diskussionen sind die Reaktionen auf die *Weimarer Ausgabe* (Johann W. von Goethe, *Goethes Werke*, hg. im Auftrage der Großherzogin Sophie von Sachsen. 144 Bde. Fotomechan. Nachdruck [Weimar 1887–1919], München 1987 [WA]), der man seit ihrem Erscheinen vorwirft, Goethes *Ausgabe letzter Hand* zu unkritisch zur Grundlage genommen zu haben, so Maria Schedl-Jokl, „Anmerkungen zur Geschichte der Weimarer Ausgabe“, in: Dies. (Hg.), *Goethes Werke*, Bd. 144: *Supplement zur Weimarer Ausgabe im Deutschen Taschenbuch-Verlag*, München 1987, 19–23, hier: 20 f.

²⁵ In den „Grundsätzen für die Weimarische Ausgabe von Goethes Werken“ (Schedl-Jokl [Anm. 24], § 16, 9) wird sogar *expressis verbis* der *Iwein*-Apparat Lachmanns als Muster vorgegeben.

und ahistorischer Textsetzung“²⁶ schwanken. Dass die Ausgaben, gerade auch jene von Texten aus der Zeit vor dem Buchdruck, jene Varianz, die das Signum einer historisch anderen Textkultur ist, in den Apparat abdrängen – er steht nicht nur unten auf der Seite oder gar im Anhang, sondern wird in kleinerer Schrift und engeren Abständen gedruckt –, ist Ausdruck dafür, dass die ahistorische Zurichtung, welche die spezifische Textualität der Überlieferung zerstört, oft Priorität hat.

Die digitale Edition unterliegt den praktischen Zwängen der Papieredition nicht mehr. Während diese im Grunde gar keine andere Wahl hat, als komplexe Überlieferungen auf einen Herausgebertext zu reduzieren, kann jene es unternehmen, die Überlieferung so vollständig wie möglich abzubilden. Die einzelnen Textzeugen werden dann nicht textkritisch miteinander ‚verrechnet‘, sondern durch Hyperlinks und Synopsen miteinander ‚verschaltet‘, wodurch sowohl ihre Verschiedenheit als auch ihr Verhältnis zueinander anschaulich gemacht werden können.²⁷ Aus der ‚universal‘ wird so die ‚universal edition[]‘,²⁸ die keinen Text im konventionellen Sinne mehr bietet, sondern eine textuelle Informationssammlung.²⁹ Der Herausgeber verliert im Zuge dieser geänderten Zielsetzungen jene ‚Stärke‘, die ihm die Textkritik verliehen hat.³⁰ Er wählt nicht aus, er rekonstruiert nicht, sondern er stellt nur noch zusammen. Dem entspricht ein weitreichender Verzicht auf textkritische Operationen auch im Detail. Die Textzeugen werden möglichst so präsentiert, wie sie (unter Ausblendung ihrer nicht reproduzierbaren Materialität) sind, oft in einer mehrschichtigen Aufbereitung, die am Digitalisat ansetzt – dieses kann seinerseits in dynamischer Weise in die Edition eingebracht werden –³¹, diesem eine diplomatische Transkription beigibt und den Text vielleicht noch behutsam bearbeitet, indem sie das Überlieferte durch Ergänzung einer modernen Interpunktion sowie durch geringfügige orthographische und/oder phonetisch-phonologische Anpassungen sowie die Korrektur handgreiflicher Fehler aufbereitet.³² Häufig aber – insbesondere bei der Dokumentation von Entwürfen und Werk(vor)stufen – fehlt dieser letzte Schritt, und man belässt es bei Digitalisat und

²⁶ Sahle (Anm. 1), Bd. 1, 132.

²⁷ Stephen G. Nichols, „Dynamic Reading of Medieval Manuscripts“, in: *Florilegium* 32 (2015), 19–57, hier: 23 f.; Peter Robinson, „Electronic Editions Which we Have Made and Which we Want to Make“, in: Ciula/Stella (Anm. 6), 1–12, hier: 1.

²⁸ Mats Dahlström, „Digital Incunables. Versionality and Versatility in Digital Scholarly Editions“, in: Peter Linde/John W. T. Smith/Elena Emelianova (Hg.), *Electronic Publishing in the Third Millennium. Proceedings of an ICC/IFIP Conference Held at the Kaliningrad State University, Kaliningrad/Svetlogorsk, Russia, 17–19 August, 2000*, Washington 2000, 224–234.

²⁹ In diese Richtung argumentiert Shillingsburg (Anm. 6).

³⁰ Vgl. Dahlström (Anm. 6), 7; Peter M. W. Robinson, „Is There a Text in These Variants?“, in: Finneran (Anm. 6), 99–115.

³¹ Gabler (Anm. 6), 49 f.

³² So etwa in Braun/Glauch/Kragl (Anm. 3).

Transkription.³³ Der Text erscheint damit in einer Gestalt, die unmittelbar an seine originäre Medialität rückgebunden ist. Während diese Zurückhaltung des Herausgebers bei modernen Texten, die – etwa über Autorhandschriften oder Ausgaben letzter Hand – gut zu greifen sind, vergleichsweise geringe Folgen für die Textgestalt zeitigt, bieten digitale Editionen bei ‚alter‘ Literatur Texte, die häufig nur noch von Spezialisten gelesen werden können.³⁴

Diesen Bruch mit den Prinzipien der ‚klassischen‘ Editorik begründet etwa die Mediävistik mit Überlegungen aus dem Umkreis der *New Philology*, die massive Zweifel an den Prämissen der Lachmann’schen Textkritik formuliert³⁵ und „geschwächten Autoren, Editoren und Lesern [...] das gestärkte Dokument“³⁶ gegenüberstellt. Auch die *Material Philology* und der *Iconic Turn* sind in diesem Zusammenhang zu nennen. Schließlich entspricht die „Tendenz zur Kontextualisierung“³⁷, die der digitalen Edition insofern innewohnt, als sie vermehrt begleitende Dokumente aufnimmt oder auf solche verlinkt, kulturwissenschaftlichen Ansätzen, wonach Texte nur in ihrem Kontext verständlich sind. Das positivistische Konzept des Kommentars wird dabei weit überschritten.³⁸ Es ist wohl kein Zufall, dass sich die Editionswissenschaft just zu der Zeit von der *Variance* und *Mouvance* sowie von der Materialität, Ikonizität und Kulturalität der handschriftlichen Überlieferung fasziniert zeigt, als sich auch der Wechsel vom Buch zur EDV vollzieht.³⁹ Damit sei nicht gesagt, dass diese avancierten Theoriebildungen reine Rückprojektionen postmoderner Befindlichkeiten darstellen. Dennoch scheint sich eine neue Harmonie einzustellen zwischen der Vorstellung, die man sich von historischer Textualität theoretisch macht, und der praktischen Herangehensweise, mit der man sie digital aufzubereiten sucht.⁴⁰ Weder ist zu

³³ Bohnenkamp, u. a. (Anm. 10); Deegan/Sutherland u. a. (Anm. 10); Girard/Leclerc (Anm. 10); Universitätsbibliothek Heidelberg/Universität Heidelberg (Anm. 10).

³⁴ Universitätsbibliothek Heidelberg/Universität Heidelberg (Anm. 10).

³⁵ Bernard Cerquiglini, *Éloge de la variante. Histoire critique de la philologie*, Paris 1989; diverse Autoren, Kapitel „The New Philology“, in: *Speculum* 65/1 (1990), 1–108; diese und verwandte kritische Stimmen zusammenfassend Sahle (Anm. 1), Bd. 1, 137–143 und 193–201.

³⁶ Sahle (Anm. 1), Bd. 1, 216.

³⁷ Sahle (Anm. 1), Bd. 2, 174.

³⁸ Gabler (Anm. 6), 53 f.

³⁹ Hans Ulrich Gumbrecht, „Philology and the Complex Present“, in: *Florilegium* 32 (2015), 273–281, hier: 278; Markus Stock, „Introduction. Philological Moves“, in: *Florilegium* 32 (2015), 1–17, hier: 3 f.; Andrew Taylor, „Getting Technology and Not Getting Theory. The New Philology after Twenty-Five Years“, in: *Florilegium* 32 (2015), 131–155, hier: 132. Dagegen weist Nichols (Anm. 21), 21 darauf hin, dass die Digitalisierung von Handschriften sich erst nach der Ausrufung der ‚Neuen Philologie‘ durchsetzte.

⁴⁰ Zu diesem Zusammenhang von Theorie und Praxis siehe Pierazzo (Anm. 12), 44 f.; Sahle (Anm. 14), 235; Francesco Stella, „Digital Philology, Medieval Texts, and the *Corpus* of Latin Rhythms, a Digital Edition of Music and Poems“, in: Ciula/Stella (Anm. 6), 223–249, hier: 228–236.

sagen, was hier zuerst ist oder war, noch ist klar zu ersehen, wo diese Harmonie eine scheinbare, wo aber eine tatsächliche ist dergestalt, dass die modernen Methoden der Aufbereitung den ‚alten‘ Verhältnissen gemäß sind. Allerdings wird man zugeben müssen, dass diese Passgenauigkeit einmal größer und einmal kleiner ist und dass damit die Vielgestaltigkeit der historischen Verhältnisse unter dem Paradigma einer umfassenden, wenig distinkten *Erschließung, Sicherung und Vergegenwärtigung unseres kulturellen Erbes*⁴¹ (so der aktuelle Titel des Akademienprogramms) abermals empfindlich homogenisiert wird.

Auch bei diesen aktuellen editionsphilologischen Entwicklungen ist der Umgang mit Textzuständen jüngerer literarhistorischer Epochen, bei denen die Überlieferung in der Regel mit Autortexten aufwarten kann, der Handhabung älterer Literatur verwandt. So gewinnt auch in den Neuphilologien die Idee der dokumentarischen bzw. archivalischen oder der Faksimile-Edition an Boden,⁴² die sich in unmittelbarer Nähe zum überlieferten Zeichenträger zu halten sucht. Solche Editionsformen werten das gesamte Material in dem Sinne auf, dass dieses nicht mehr zu einem kritischen Apparat kondensiert, sondern in seiner Gesamtheit zur Verfügung gestellt wird, und dies meist, ohne dass die Relevanz der einzelnen Textzeugen irgendwie deutlich gemacht würde, weil eine solche Wertung ja einem ‚Rückfall‘ in frühere Gepflogenheiten gleichkäme. Der Vergleich des *Faust II* in der *Weimarer Ausgabe* mit dem *Faust II* der Online-Edition ist dafür ein besonders markantes Beispiel.

1.3 Rolle des Rezipienten

Gedruckte Buchliteratur legt eine bestimmte Lesehaltung nahe: die lineare, vollständige und konzentrierte Lektüre eines gesicherten, fixierten Textes. Der *Faust II* der *Weimarer Ausgabe* ist dazu gedacht, von vorne bis hinten gelesen zu werden, wenn dies auch bei einem Apparat dieser Fülle eine extreme Anforderung an Aufmerksamkeit und Muße des (sicher nur wissenschaftlichen) Lesers bedeutet. Jede Buchedition kann sich also nur an einen bestimmten Lesertyp wenden, etwa an Literaturliebhaber oder an Wissenschaftler, und Letztere haben sehr unterschiedliche, da spezialistische Interessen. Um den daraus resultierenden Zielkonflikt zu lösen, entschied man sich einerseits, die historisch-kritische Ausgabe zur Norm zu erklären, und andererseits, von ihr Derivate für breitere Lesergruppen wie Lese- und Studienausgaben abzuleiten. Auch wurden Entscheidungen der Editoren, die medialen Gegebenheiten geschuldet waren, mit Rücksicht auf

⁴¹ <https://www.forschung.europa-uni.de/en/ausschreibung/13692> (letzter Aufruf 22.09.2017).

⁴² Gabler (Anm. 6), 51. Dass die archivalische Funktion quasi von Beginn an zu den definitiven Merkmalen der digitalen Edition gehört hat, konstatiert John Lavagnino, „Electronic Editions and the Needs of Readers“, in: W. Speed Hill (Hg.), *New Ways of Looking at Old Texts*, Bd. II: *Papers of the Renaissance English Text Society, 1992–1996*. Tempe, Arizona 1998, 149–156, hier: 149.

den Leser begründet, den man nicht mit unübersichtlichen Synopsen oder überbordenden Apparaten behelligen könne. Auf ähnliche Weise werden regulierende, normierende und modernisierende Eingriffe mit Lesererwartungen gerechtfertigt. Das mag einer Ausrede gleichkommen, verweist aber eben auch auf die Ökonomie von Lektürepraktiken. Insgesamt ist der Leser einer kritischen Edition ‚schwach‘, weil er die Arbeit des Herausgebers letztlich nur hinterfragen kann, indem er sich selbst den Quellen zuwendet.

Im digitalen Medium kann die Tätigkeit des Editors in ganz anderer Weise transparent werden.⁴³ Statt sich Autoritäten (Herausgeber, Verlag, Rezensent) anzuvertrauen, kann sich jeder Rezipient im Rückgriff auf das ungleich reichhaltigere, in der Edition selbst enthaltene Quellenmaterial sein eigenes, begründetes Urteil über die Qualität der Edition, ihre Zuverlässigkeit und ihre Prinzipien bilden.⁴⁴ Auch indem er im individuellen Umgang mit der zur Verfügung gestellten Textmenge latent in die alte Herausgeberrolle schlüpft, gewinnt der Rezipient an ‚Stärke‘.⁴⁵ Freiheit setzt freilich mündige Subjekte voraus, und so gibt die digitale Edition dem Rezipienten nicht nur die Gelegenheit, eigene Entscheidungen zu treffen, sondern sie zwingt ihn geradezu, dies zu tun. Nicht nur muss er aus dem gebotenen Material eine eigene Auswahl treffen, sondern er hat sich auch für die Gestalt zu entscheiden, in der ihm dieses gegenüberstehen soll.⁴⁶ Damit wird aus dem Leser, der sich in ein Werk versenkt und es zu verstehen versucht, unweigerlich ein Benutzer, der sich dieses in Auseinandersetzung mit der Textgeschichte und dem Kontext erarbeitet.⁴⁷ Um das zu tun, muss sich der Benutzer mit den Werkzeugen vertraut machen, die das Navigieren im Textarchiv erst ermöglichen, vor allem aber wird ihm ein eigenes philologisches Urteil abverlangt.

Im Extremfall kann die gebotene Vielfalt der Versionen dazu führen, dass die Benutzer gar keinen Zugang zum Material finden. Die Edition des *Welschen Gasts* antwortet mit einem „Schnelleinstieg“⁴⁸ auf dieses Problem, doch funktioniert dieser am besten für wissenschaftliche Experten, die bereits wissen, wo sie hinfahren wollen, und sich nicht erst orientieren müssen. Das führt dann zu einer paradoxen

⁴³Thomas Stäcker, „Creating the Knowledge Site. Elektronische Editionen als Aufgabe einer Forschungsbibliothek“, in: Christiane u. a. (Hg.), *Digitale Edition und Forschungsbibliothek. Beiträge der Fachtagung im Philosophicum der Universität Mainz am 13. und 14. Januar 2011*, Wiesbaden 2011, 107–126, hier: 113.

⁴⁴Robinson (Anm. 27), 1.

⁴⁵Vgl. Charles L. Ross, „The Electronic Text and the Death of the Critical Edition“, in: Finneran (Anm. 6), 225–231.

⁴⁶Pierazzo (Anm. 12), 51 f.

⁴⁷Diese Unterscheidung von Leser und Benutzer nach Krista S. G. Rasmussen, „Reading or Using a Digital Edition? Reader Roles in Scholarly Editions“, in: Driscoll/Pierazzo (Anm. 6), 119–133, hier: 127, die sie freilich beide als unterschiedliche Typen von Rezipienten digitaler Editionen einführt. Letztere lassen aber voraussetzungsloses Lesen im Grunde gar nicht mehr zu.

⁴⁸Universitätsbibliothek Heidelberg/Universität Heidelberg (Anm. 10).

Situation: Frei im Netz erreichbare Editionen – überall, jederzeit, sofort und umsonst auf eine Edition zugreifen zu können, entspricht einer Erwartung, die unseren Umgang mit Texten allgemein zunehmend prägt – erweitern den Benutzerkreis weit über den der Buchkäufer und Bibliotheksbenutzer hinaus, und zwar nicht nur ihrem Anspruch nach, sondern auch in der Realität, wie sich an Klickzahlen digitaler Editionen ablesen lässt, die die der Verkäufe und Ausleihen gedruckter Editionen weit übertreffen.⁴⁹ Und indem digitale Editionen verschiedene Versionen der Texte anbieten, suchen sie auch konzeptionell unterschiedlichen Bedürfnissen gerecht zu werden – und werden gerade dadurch so komplex, dass sie ein allgemeines Publikum latent überfordern.

Besonders dem wissenschaftlichen Benutzer erlegt die digitale Edition Verpflichtungen auf wie die, sich mit dem handschriftlich überlieferten Wortlaut zu befassen, sich in die Genese eines Werks einzuarbeiten oder die Parallelüberlieferung vollständig zur Kenntnis zu nehmen. Niemand wird sich in solchen Fragen mehr auf die Unzulänglichkeiten der verwendeten Ausgabe herausreden können, und man wird zugeben müssen, dass diese Ausrede in arbeitspragmatischer Hinsicht oft auch etwas Entlastendes hatte. Die Entlastung hat freilich unterschiedlichen Charakter, wie das Beispiel der varianten Überlieferung verdeutlicht: Ist die Varianz so gering wie die zwischen den Zeugen des *Faust I*, bewahrt ein Lesetext vor der Enttäuschung, die sich rasch einstellt, wenn man vorgeblich variante Verse anklickt, sie eingehender studiert und dann zumeist doch nur orthographische Varianten findet. (Auch für Letztere mag es ein wissenschaftliches Interesse geben, aber es ist hoch spezialisiert und sicher nicht das der meisten literaturwissenschaftlichen Leser.) Bei einem mehrfach überlieferten Minnelied dürfte das Interesse des Interpreten daran, die Überlieferungsvarianten übersichtlich dargeboten zu bekommen, die Entlastung egalisieren, welche die Komplexitätsreduktion durch eine kritische Edition bedeutet. Aber was macht der Wissenschaftler, der eine Geschichte der mittelhochdeutschen Literatur schreiben und sich dazu auch den Minnesang ansehen möchte? Wie findet er die ‚Abkürzung‘, die er braucht? Und wie beruhigt er das schlechte Gewissen, das der Philologe in ihm haben wird, wenn er sie nutzt?

Nicht verzichtbar scheint die Entlastung durch kritische Eingriffe bei umfangreichen, reich überlieferten oder bei sehr schlecht überlieferten Texten. Die digitale *Parzival*-Ausgabe wird den an die 25.000 Verse langen Roman nach allen 15 vollständigen Handschriften, einem Frühdruck sowie 71 Fragmenten herausgeben.⁵⁰ Wie sie benutzt wird, muss sich noch erweisen; dass der Herausgeber selbst bereits über eine Art Studienausgabe nachdenkt,⁵¹ die der souverän

⁴⁹Price (Anm. 6), 446 f.

⁵⁰Diese Angaben nach Michael Stolz, „Benutzerführung in digitalen Editionen. Erfahrungen aus dem *Parzival*-Projekt“, in: Fritze u. a. (Anm. 43), 49–80, hier: 55.

⁵¹Michael Stolz, „Von den Fassungen zur Eintextedition. Eine neue Leseausgabe von Wolframs *Parzival*“, in: Dorothea Klein (Hg.), *Überlieferungsgeschichte transdisziplinär. Neue Perspektiven auf ein germanistisches Forschungsparadigma*, Wiesbaden 2016, 353–388.

selezierenden Edition Lachmanns im Ergebnis vermutlich gar nicht so fern stünde, deutet das Problem aber bereits an. Vom griechischen Neuen Testament gibt es gar derart viele Zeugen, dass es unmöglich erscheint, diese synoptisch zu edieren bzw. zu rezipieren.⁵² Ohne wertenden Textvergleich ist einer solchen Überlieferungsmasse nur stichprobenartig beizukommen, was im Falle eines auratischen Textes aber auch keine Lösung ist. Petrons *Satyricon* wiederum ist zwar nur spärlich, dafür aber in teils sehr kurzen, vielfach verderbten Fragmenten erhalten,⁵³ sodass man ohne die rekonstruierenden Bemühungen der Herausgeber vor einem Trümmerhaufen stünde, der sich unmöglich (oder nur nach langer intensiver Beschäftigung) gedanklich fassen ließe.

Wechselt man vom einzelnen Benutzer auf die Ebene der Benutzergemeinschaft, dürfte ein weiterer Effekt digitaler Editionen darin bestehen, dass deren Fluidität – hervorgerufen durch das Nebeneinander der Fassungen, die Alternativen der Darstellungsoptionen und das Nacheinander der veröffentlichten Versionen – die wissenschaftliche Community um einen gemeinsamen Bezugspunkt bringt.⁵⁴ Gerade für Disziplinen, die aufgrund unterschiedlicher, mitunter geradewegs inkompatibler theoretischer Annahmen ohnehin in Gruppen auseinanderzufallen drohen, ist diese Gefahr nicht zu unterschätzen.

Für eine Übergangszeit mögen alle diese Schwierigkeiten, die sich aus der fehlenden Linearität und mangelnden Festigkeit digitaler Editionen ergeben, noch dadurch aufgefangen werden, dass die Benutzer die Texte bereits aus älteren gedruckten Editionen kennen und das Wissen, das ihnen diese vermittelt haben, in das neue Medium mitnehmen. Für die digital sozialisierten Leser, die keinen solchen Hintergrund mehr haben, werden sie sich zuspitzen. Überzeugende Lösungen für die entsprechenden Probleme müssen erst noch gefunden werden. Den neuen Ausgaben die alten Texte der gedruckten Bücher als Digitalisat beizugeben, kann nicht der Weisheit letzter Schluss sein. Auch die Hybridedition, die dem elektronisch verfügbaren vollständigen Material einen gedruckten Lesetext an die Seite stellt, löst die Probleme bei der Rezeption der digitalen Edition gerade nicht.⁵⁵

⁵² Kurt Aland/Barbara Aland, *Der Text des Neuen Testaments. Einführung in die wissenschaftlichen Ausgaben sowie in Theorie und Praxis der modernen Textkritik*, Stuttgart ²1989 [1982]; Bruce M. Metzger, *The Text of the New Testament. Its Transmission, Corruption, and Restoration*, New York/Oxford ³1992 [1964].

⁵³ Prägnant zusammenfassend die Einleitung der Loeb-Ausgabe: *Petronius*. With an English Translation by Michael Heseltine. *Seneca: Apocolocyntosis*. With an English Translation by W. H. D. Rouse, hg. von E. H. Warmington, Cambridge/London 1969, xix–xxx.

⁵⁴ Pierazzo (Anm. 12), 56 f.

⁵⁵ Angekündigt sind Druckausgaben für die *Historisch-kritische Faustedition* (Bohnenkamp u. a. [Anm. 10]) sowie, als synoptische Fassungsedition, für die *Parzival*-Edition, vgl. Stolz (Anm. 51). Als Konzept wird die Hybridedition in einer Phase vertreten, als sich die Möglichkeiten der digitalen Edition erst abzuzeichnen begannen, etwa bei John Bryant, *The Fluid Text. A Theory of Revision and Editing for Book and Screen*, Ann Arbor 2002, 149–151.

Wie sähe demgegenüber eine Lektürepraxis aus, auf die die umfassende Darbietung der Überlieferung samt ihrer theoretischen Unterfütterung hinführt? Zum einen tritt der Editionstext nicht mehr an die Stelle des historischen Dokuments, sondern – da dieses als Abbildung gegenwärtig ist – an dessen Seite; und er steht nicht mehr für sich, vielmehr ist er eingebettet in ein Gefüge von Texten. Das nimmt ihm einen Teil seiner Autorität, während die Materialität und Visualität des jeweiligen Zeugen stärker ins Bewusstsein treten. Zum anderen ist die digitale Edition von vornherein nicht daraufhin angelegt, im herkömmlichen Sinne gelesen zu werden; sie begünstigt und fordert andere, neue Formen der Benutzung, die sich aus ihren Prämissen herleiten: An die Stelle der vollständigen tritt – schon aufgrund der Datenmenge – die selektive Lektüre, an die Stelle der Konzentration auf den einen autorisierten Text das Interesse an der Partikularität des einzelnen Überlieferungsträgers, dessen Fassung in der Synopse den anderen Fassungen gegenübersteht, an die Stelle der Linearität des Lektüreprozesses ein von Textsuchen und Hypertextstrukturen geleitetes Springen von Stelle zu Stelle,⁵⁶ das zu „heuristische[n] Strategien des ‚Stöbers‘“⁵⁷ führt, an die Stelle des Wort-für-Wort-Lesens eine (im Prinzip) statistische Auswertung des Textmaterials. ‚Browsen statt Lesen‘, könnte man diese Tendenzen auf den Begriff bringen.⁵⁸ Solche Veränderungen der wissenschaftlichen Lektürepraxis sind Teil eines gesamtgesellschaftlichen Veränderungsprozesses im Umgang mit Texten, den die digitalen Medien angestoßen haben.⁵⁹ Wer digital liest, liest anders,⁶⁰ und zwar im alltäglichen wie im wissenschaftlichen Rahmen.

⁵⁶Bertrand Gervais zum besonderen Zeichenstatus des Hyperlinks und dessen Auswirkungen auf das Lesen, das passiver wird, da die Entdeckung nicht mehr durch das eigene Vorwärtsdenken in der Lektüre erfolgt: „Is There a Text on This Screen? Reading in an Era of Hypertextuality“, in: Siemens/Schreibman (Anm. 6), 183–202, hier: 197–200.

⁵⁷Sahle (Anm. 1), Bd. 2, 271.

⁵⁸Neben *Browsing* werden auch noch *Surfing* und *Navigating* als Begriffe für den Wissenserwerb im Internet diskutiert, der zwar auch auf dem Lesen beruht, sich von der Benutzung eines Buchs aber eben doch signifikant unterscheidet, vgl. Gervais (Anm. 56), 186.

⁵⁹Einen aktuellen Überblick bietet Ursula Rautenberg/Ute Schneider (Hg.), *Lesen. Ein interdisziplinäres Handbuch*, Berlin/Boston 2015, darin die Abschn. 2.2.4 „Digitale Lesemedien“ (Axel Kuhn/Svenja Hagenhoff, 361–380), 3.3.4 „Nutzergenerierte Texte in digitalen Netzwerken“ (Dies., 679–699) und 4.1.5 „Lesen und Medien in der zweiten Hälfte des 20. Jahrhunderts“ (Hans-Dieter Kübler, 793–812). Genannt werden u. a. Änderungen der Aufmerksamkeitsspanne, sprunghaftes, fragmentarisches und selektives Leseverhalten, rasche Wechsel des Lesetextes, dynamische Anpassungen des Lektüreobjekts (Anordnung, Typographie etc.) an die individuellen Lesegewohnheiten, intensive Nutzung von interaktiven Funktionen, generell eine gesteigerte Ungeduld des Lesers, die von der simultanen Verfügbarkeit vieler Texte (nur ein Klick) gesteigert wird (373–377). Eine intensive Interaktion zwischen Textproduzenten und -rezipienten ist bei nutzergenerierten Texten (z. B. Wiki-Systemen) zu beobachten, die, nicht zuletzt aufgrund ihrer produktionsbedingten Heterogenität, eine kritisch-selektive Rezeptionsweise geradezu einfordern (692–695). Vgl. außerdem Gervais (Anm. 56), und Christian Vandendorpe, „Reading on Screen. The New Media Sphere“, in: Siemens/Schreibman (Anm. 6), 203–215.

⁶⁰Zusammenfassend Kübler, „Lesen und Medien in der zweiten Hälfte des 20. Jahrhunderts“, in: Rautenberg/Schneider (Anm. 59), 807 f.

2 Zukunftsperspektiven der digitalen Edition

Während der Herausgeber der Zukunft seine traditionelle textkritische Rolle einbüßen wird, werden ihm neue Aufgaben zuwachsen und neue Fertigkeiten abverlangt. Vor allem spielen technische Kenntnisse eine sehr viel größere Rolle bei der editorischen Arbeit – so sind beispielsweise Probleme wie die der Zeichencodierung, der formalisierten Adressierung via PID oder der Versionierung zu lösen –, und der Editor der Zukunft wird diese entweder selbst besitzen und/oder in hohem Maße zu kollaborativem Arbeiten fähig sein müssen.⁶¹ In vielen Fällen wird er sich bei der Umsetzung seiner philologischen Wünsche freilich den konkret und meist kontingent vorhandenen technischen Möglichkeiten zu fügen haben.⁶² Der Stand der Technik bleibt das Nadelöhr für die Editorik, wenn es auch den Anschein hat, als ob die technischen Restriktionen mehr und mehr im Schwinden begriffen sind. Gerade deshalb wäre es wünschenswert, dass es in bestimmten Bereichen auch zu Standardisierungen kommt.⁶³ Bislang entwickelt beispielsweise jedes Editionsprojekt seine Präsentationstechnik (Webinterface) neu, selbst wenn es nur einzelne Texte oder kleine Korpora herausgibt. Entsprechend wirken viele Editionen wie Prototypen,⁶⁴ die nie in Serie gehen. Das neue Medium wirft aber nicht nur technische Fragen auf, sondern etwa auch urheberrechtliche oder wissenschaftssoziologische. Erstere betreffen etwa solche der Rechte an Bildern oder der Nachnutzung,⁶⁵ Letztere die der Geltung digitaler editorischer Arbeit im Wissenschaftsbetrieb.⁶⁶

⁶¹ Kollaboratives Arbeiten meint hier dreierlei: die Zusammenarbeit mit anderen Geisteswissenschaftlern, die Kooperation mit IT-Fachleuten und den Austausch mit anderen Experten auf dem Gebiet der digitalen Edition; zu den entsprechenden Gruppen vgl. Daniel Paul O'Donnell, „Disciplinary Impact and Technological Obsolescence in Digital Medieval Studies“, in: Siemens/Schreibman (Anm. 6), 65–81, hier: 72 f., und Stäcker (Anm. 43), 117–119.

⁶² Die Kontingenz ergibt sich weniger aus dem eingeplanten und bewilligten Budget für Datenmodellierung und Interfacedesign, sondern vielmehr aus der Wahl eines Technik-Partners – heute in der Praxis oft IT- oder DH-Abteilungen bzw. -Zentren von Akademien oder Universitäten. Deren sehr unterschiedliche Standards und Präferenzen für bestimmte Techniken müssen von ihren philologischen Klienten schon deshalb hingenommen werden, weil diese sie kaum je fundiert beurteilen können.

⁶³ Dazu Elmar Mittler/Malte Rehbein, Edition und Forschungsbibliothek. Chancen und Herausforderungen einer traditionsreichen Partnerschaft im digitalen Zeitalter, in: Fritze u. a. (Anm. 43), 9–21, hier: 18–20.

⁶⁴ Z. B. *Digitale Edition des steirisch-landesfürstlichen Marchfutterurbars von 1414/1426. Testversion*, bearbeitet von Matthias P. Perstling, <http://hfi.uni-graz.at/mfu/#fo674> (letzter Aufruf 19.8.2017); Reinhold Kaiser (Hg.), *Die Glossen Ekkeharts IV. im Codex Sangallensis 621. Edition (Betaversion)*, <http://orosius.monumenta.ch/> (letzter Aufruf 19.8.2017).

⁶⁵ Stäcker (Anm. 43), 121 f.

⁶⁶ Verschiedene dieser Aspekte diskutiert Ingo H. Kropač, „Work in Progress. Vom Digitalisat zum edierten Text“, in: Thumser/Tandeki (Anm. 9), 167–183, hier: 177–179.

Der Herausgeber einer digitalen Ausgabe wird künftig weniger einen einzelnen Editionstext erarbeiten, sondern vielmehr als Organisator einer umfangreichen und entsprechend komplexen Materialpräsentation hervortreten. Als „corpus editor“ hält er dann „a middle ground between the algorithm-heavy, knowledge-light approaches of computer science and the wholly manual practices of traditional editing“.⁶⁷ Er hat sicherzustellen, dass die neue Mehrdimensionalität der digitalen Edition – etwa hinsichtlich der Parallelität der Zeugen, der Einbindung von Kontextmaterial, der Anreicherung des Grundtextes mit weiteren Schichten etc. – EDV-technisch bewältigt wird, und er hat dafür Sorge zu tragen, dass der mehrdimensionale Textraum der digitalen Edition auf eine Weise präsentiert wird, die menschlichen Benutzern hinreichend Orientierung bietet (*Usability*).⁶⁸ Die Mehrdimensionalität der Textpräsentation impliziert auch eine Flexibilisierung des Arbeits- und Publikationsprozesses, weil es nunmehr nicht wesentlich ist, welcher Textzeuge zuerst bearbeitet wird, und ob zuerst ein Digitalisat oder eine Transkription, ob nur das schiere Textmaterial, z. B. in XML/TEI-Codierung, oder zugleich schon avanciertere Formen der Darstellung und Auswertung (Hyperlinks, Synopsen etc.) bereitgestellt werden. An die Stelle überkommener philologischer Maximen wie Vollständigkeit, Verlässlichkeit und Abgeschlossenheit treten die digitalen Grundsätze des allmählichen Wachsens, der augenblicklichen Aktualität, der Möglichkeit steten Verbesserns und Erweiterns.⁶⁹ Mithin findet eine Verschiebung vom Produkt zum Prozess statt.⁷⁰

Abgeschlossene Editionen wiederum werden auf Dauer nur benutzbar bleiben, wenn eine institutionelle Lösung gefunden wird, die die Langfristarchivierung gewährleistet – und zwar nicht nur im Sinne einer sicheren Speicherung der Daten, sondern auch in dem einer nachhaltigen Betreuung ihrer Präsentation.⁷¹ Denn

⁶⁷ Gregory Crane/David Bamman/Alison Jones, „ePhilology. When the Books Talk to Their Readers“, in: Siemens/Schreibman (Anm. 6), 29–64, hier: 52.

⁶⁸ Überlegungen zur Benutzerführung am Beispiel der digitalen *Parzival*-Edition bei Stolz (Anm. 50).

⁶⁹ Kropač (Anm. 66), 171 f.; Pierazzo (Anm. 12), 51; Price (Anm. 6), 441. Entsprechend sind etwa die Projekte *Historisch-kritische Faustedition* (hg. von Bohnenkamp u. a. [Anm. 10]) und *Welscher Gast digital* (hg. von Universitätsbibliothek Heidelberg/Universität Heidelberg [Anm. 10]) mit einer Betaversion an den Start gegangen.

⁷⁰ Patrick Sahle, „What is a Scholarly Digital Edition?“, in: Driscoll/Pierazzo (Anm. 6), 19–39, hier: 29.

⁷¹ Hierzu und zum Folgenden aus mediävistischer Perspektive O'Donnell (Anm. 61), 65–69. Die ‚Nachhaltigkeit‘ des XML-Formats hilft hier nicht weiter, weil dieses kein Präsentationsformat ist. Die für die Ausgabe von XML-Dateien entwickelte *Extensible Stylesheet Language Transformation (XSLT)* ist nicht so weit gediehen, dass man eine wissenschaftliche Edition allein auf ihrer Grundlage publizieren könnte (Sahle [Anm. 1], Bd. 2, 74). Entsprechend kann auch das TextGrid-Repository, <https://textgridrep.org> (letzter Aufruf 19.8.2017) nur für die langfristige Verfügbarkeit der Rohdaten als Dateien, nicht aber der Edition selbst sorgen; ganz abgesehen davon, dass dieses Dateiarhiv nichts anderes als eine ‚öffentliche Festplatte‘ darstellt, deren Befüllung im Belieben der Teilnehmer steht.

die Möglichkeit zu deren ständiger Weiterentwicklung verkehrt sich schon nach wenigen Jahren in den Nachteil, sie weiterentwickeln zu müssen, wenn nicht entweder die technische Lauffähigkeit oder der medial-ästhetische und benutzungspragmatische *State of the Art* verloren gegeben werden sollen. Wenn für diese Weiterentwicklung irgendwann keine Ressourcen mehr bereitstehen, bedeutet ein von einem technischen Entwicklungssprung geforderter Formatwechsel bislang oft ein Einfrieren des Projekts.⁷² Die Technik, die zunächst eine Befreiung des Herausgebers von den Beschränkungen der Buchedition verspricht,⁷³ stellt sich so nach einigen Jahren als Hindernis heraus.⁷⁴ Es kann nur aus dem Weg geräumt werden, wenn öffentliche Einrichtungen wie Bibliotheken oder (größtenteils noch zu errichtende) (supra-)nationale Infrastrukturen die digitalen Editionen in ihre Obhut nehmen und sie langfristig am Laufen halten;⁷⁵ alles andere wäre eine nicht zu verantwortende Vergeudung der Ressourcen, die in die Erarbeitung der digitalen Editionen geflossen sind.

Des Weiteren bedarf es, sollen die Möglichkeiten der digitalen Edition ausgeschöpft werden, eines Einstellungswandels aufseiten der Editoren wie aufseiten der Benutzer. Denn noch finden sich verschiedene (im Wort- und im konventionellen Sinne) ‚reaktionäre‘ Abwehrmechanismen gegen die Eigenlogik der digitalen Edition.⁷⁶ Sie treten – wenn man von einer Totalverweigerung gegen alles Digitale absieht (auch die gibt es) – zutage als Schwierigkeiten in deren Handhabung, die sich immer dann einstellen, wenn sich die Beteiligten nicht zur Gänze von traditionellen Vorstellungen freimachen können. Das fängt mit Kleinigkeiten wie der an, dass etwa Apparate nach wie vor unter dem Text stehen und damit abgewertet werden, was dem alten, nicht dem neuen Medium entspricht.⁷⁷

⁷² Es lohnt nicht, diese verlassenen Projekte im Einzelnen aufzuführen. Man findet sie rasch, wenn man sich durch die Liste von Sahle (Anm. 10), klickt.

⁷³ Diese Perspektive etwa bei H. T. M. van Vliet, „Die elektronische Edition als Lösung alter editorischer Probleme? Beobachtungen eines Editors“, in: Václav Bok/Frank Shaw (Hg.), *Magister et amicus. Fs. für Kurt Gärtner zum 65. Geburtstag*, o. O. 2003, 269–288.

⁷⁴ Vgl. <http://ebeowulf.uky.edu/#> (letzter Aufruf 19.8.2017) mit dem Bericht über den erzwungenen Umstieg zwischen 2013 und 2015 von DVD/Java auf Online/Javascript. Dieses 1994 begonnene Projekt, das den *Electronic Beowulf* jetzt in der Version 4.0 präsentiert (hg. von Kevin Kiernan), ist eines der wenigen, die den Kampf gegen die technische Obsoleszenz nicht irgendwann aufgegeben haben.

⁷⁵ Überlegungen hierzu bei Reinhard Altenhöner, „Trau, schau, wem. Zur Authentizität und Langzeitverfügbarkeit digitaler Objekte“, in: Fritze u. a. (Anm. 43), 153–170; Peter Boot/Joris van Zundert, „The Digital Edition 2.0 and The Digital Library. Services, not Resources“, in: Fritze u. a. (Anm. 43), 141–152, hier: 145–152; zur technischen Seite des Problems: Alexander Czmiel, *Sustainable Publishing. Standardization Possibilities for Digital Scholarly Edition Technology*, <http://dixit.uni-koeln.de/convention-2-abstracts/#czmiel> (letzter Aufruf 19.8.2017).

⁷⁶ Hierzu und zum Folgenden Joris van Zundert, „Barely Beyond the Book?“, in: Driscoll/Pierazzo (Anm. 6), 83–106, hier: 85 f.

⁷⁷ Gabler (Anm. 6), 49.

(Auch unsere eigene Edition *Lyrik des deutschen Mittelalters* bewahrt z. B. das Layout des gedruckten Buchs in diesem und weiteren Punkten.) Dem digitalen Medium wären aber vor allem größere Textmengen angemessen, denn nur hier kann die EDV ihre ganze Stärke ausspielen. Insofern kommt es im Grunde einer latenten Verweigerungshaltung hinsichtlich der digitalen Möglichkeiten gleich, wenn man als Herausgeber mit hohem Aufwand nach wie vor nur kleine Korpora bearbeitet, an der Maxime der Fehlerfreiheit festhält und deshalb etwa mit hohem Aufwand OCRs korrigiert, wo doch eine Big-Data-Logik gerade darauf baut, dass auch das Ungefähre auf das Wesentliche deutet. In der Folge verschwimmen zumindest die Grenzen zwischen Editions- und Digitalisierungsprojekten.⁷⁸ Begriffe wie „critical digitization“⁷⁹ versuchen, diese Übergangszone zu vermessen. Auch in diesem Sinne wäre – wie schon oben angedeutet – zu überlegen, ob die Wendung ‚digitale Edition‘ nicht per se paradoxal ist.

Benutzerseitig äußert sich das Festhalten am Gewohnten beispielsweise in Strategien, die Textmengen digitaler Editionen irgendwie auf eine Lesefassung hin zu reduzieren, oder in der Tendenz, digitale Editionen nicht am Bildschirm zu lesen, sondern sie als Datenmine für Papiausdrucke zu verwenden. Ein solch partielles Festhalten an Konventionen ist in seiner paradoxen Natur typisch für einen Medienwandel, und es ist charakteristisch, dass die ‚erste Generation‘ digitaler Editoren und Leser diesen Paradoxien in besonderer Weise unterliegt. Im Moment will die digitale Edition zugleich ebenso die Optimierung der Buchedition sein wie deren Transformation in das neue Medium, dessen Möglichkeiten sich ja auch erst nach und nach abzeichnen. Auf mittlere Sicht wird sich mutmaßlich dessen Logik durchsetzen und zu einem neuen System führen, das die ‚reaktionären‘ Strategien obsolet macht. Dessen Umrisse könnten die folgenden Punkte abstecken, die die bisher umrissenen Tendenzen zugleich resümierend aufgreifen und in letzter Konsequenz weiterdenken:

⁷⁸Thomas Stäcker, „Konzepte zur Bereitstellung digitalisierter frühneuzeitlicher Quellen“, in: Rainer Hering u. a. (Hg.), *Forschung in der digitalen Welt. Sicherung, Erschließung und Aufbereitung von Wissensbeständen. Tagung des Staatsarchivs Hamburg und des Zentrums Geisteswissenschaften in der digitalen Welt an der Universität Hamburg am 10. und 11. April 2006*, Hamburg 2006, 143–152, hier: 151, der als Mindestanforderung für eine Edition Digitalisat und Transkription nennt, zugleich aber seine Unsicherheit hierüber eingesteht. Ähnlich konstatiert schon 2001 Patrick Sahle, „Die Grenzen des Begriffs ‚Edition‘ verschwimmen“, siehe: Ders., „Digitale Editionstechniken und historische Quellen“, in: Stuart Jenks/Stephanie Marra (Hg.), *Internet-Handbuch Geschichte*, Köln/Weimar/Wien 2001, 153–166, hier: 161. Ausführlicher hierzu Patrick Sahle, „Digitales Archiv – Digitale Edition. Anmerkungen zur Begriffsklärung“, in: Michael Stolz/Lucas Marco Gisi/Jan Loop (Hg.), *Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien. Eine Standortbestimmung*, Zürich 2007, 64–84.

⁷⁹Mats Dahlström, „Editing Libraries“, in: Fritze u. a. (Anm. 43), 91–106, hier: 95. Genauer zum Konzept und dessen Verhältnis zur Massendigitalisierung einerseits und zur wissenschaftlichen Edition andererseits ebd., 95–103.

1. Möglich erscheinen Formen einer „personalisierte[n]‘ Benutzung“, etwa „Kommentar- und Notizfunktionen oder die Speicherung von Pfaden, Auswahlen, Trefferlisten“.⁸⁰ Der nächste Schritt auf diesem Weg wäre der Einsatz von Künstlicher Intelligenz, der es erlauben würde, die Texte bzw. deren Darbietung automatisch an den jeweiligen Benutzer anzupassen.⁸¹ Eine solche algorithmische Entlastung des Rezipienten ist zwar hochproblematisch, weil sie leicht zur Entmündigung führen kann, dennoch mag sie angesichts eines großen, durch vieldimensionale Vernetzung labyrinthisch wirkenden Datenangebots das kleinere Übel sein.
2. Die Texte digitaler Editionen werden mit Daten angereichert, sie werden etwa syntaktisch analysiert (*Part-of-Speech Tagging*), lemmatisiert (flektierte Wörter werden auf ihre Grundform zurückgeführt) oder anderweitig annotiert (Entitäten, Zeit, Raum, Institutionen und ähnliche werden ausgezeichnet und gegebenenfalls mit *Identifiern* versehen).⁸² Diese Informationen könnten über Schnittstellen wie *Linked Open Data* editionenübergreifend ‚geerntet‘, in großen Katalogen gesammelt und zur Vernetzung verschiedener Projekte genutzt werden. Die Grenze zwischen dem ‚Inhalt‘ und der ‚Umwelt‘ eines derart ausgebauten Textangebots wird in der digitalen Sphäre durchlässig.
3. Auch die Benutzer versehen die Texte mit weiteren Daten und bieten ihre Annotationen – Stichwort: Web 2.0 bzw. *Crowdsourcing* – zur Weiternutzung über die Editionsplattform an.⁸³ Auch machen sie dort zusätzliches Kontextmaterial oder eigene Kommentare zugänglich. Im Effekt bedeutet das: „Die Grenze zwischen dem Editor und dem Leser beginnt sich aufzulösen.“⁸⁴ Ihr Verhältnis wird demokratisiert, indem der Benutzer zum Co-Editor nicht nur im übertragenen, sondern auch im wörtlichen Sinne wird. Damit sich solche Ansätze entwickeln, sind Fragen der Qualitätskontrolle genauso zu klären wie solche der Reputationszuweisung. Einer ihrer Effekte wäre es, dass Editionen nie abgeschlossen sind, sondern sich weiterentwickeln, solange sie genutzt werden.
4. Eine durch diverse Auszeichnungen vorbereitete Edition dient als Ressource für *Text Mining*, für formalisierte Zugriffe, die vor allem die Lexik bzw. Semantik des Textes in den Blick bekommen. Die Edition erscheint dann als Textdatenbank, die es möglich macht, weiterreichende Transformationen bzw. Analysen vorzunehmen, welche die Texte in Bündel von Visualisierungsdiagrammen verwandeln. Diese können neue Fragen beantworten, aber auch alte

⁸⁰ Sahle (Anm. 1), Bd. 2, 258.

⁸¹ Als Vision formuliert bei Crane/Bamman/Jones (Anm. 67), 45 f.

⁸² Zum *Part-of-Speech Tagging* und zur Lemmatisierung siehe Stäcker (Anm. 43), 115; zu Tendenzen, Editionen durch Annotationen anzureichern siehe Price (Anm. 6), 435.

⁸³ Hierzu und zum Folgenden Boot/Zundert (Anm. 75), 143 f.; O’Donnell (Anm. 61), 77; Price (Anm. 6), 447; Rasmussen (Anm. 47), 127 f.; Robinson (Anm. 27), 10 f.; Sahle (Anm. 70), 30; Stäcker (Anm. 43), 118 f.

⁸⁴ Sahle (Anm. 1), Bd. 2, 258.

philologische wie die nach dem Stemma, indem dieses – eine entsprechende Manuskriptkultur vorausgesetzt – mit bioinformatischen Methoden errechnet und graphisch dargestellt wird.⁸⁵

5. Die Editionen enthalten den Werkzeugkasten für die Textanalyse unmittelbar selbst.⁸⁶ Das hat den Vorteil, dass auch technisch weniger versierte Benutzer, die mit den entsprechenden Repositorien nicht vertraut sind, die Analysetools – etwa solche zur automatisierten Analyse der Metrik, der Varianz, der Worthäufigkeit und -frequenz oder der Stilometrie – leichter einsetzen können. Denn diese sind dann auf die Datenformate, aber auch andere Eigenheiten der Editionstexte abgestimmt – etwa auf die der jeweiligen historischen Sprachstufe –, und sie sind von vorneherein auf die Bedienung durch Literaturwissenschaftler (und eben nicht Linguisten) ausgelegt. Damit ein solcher Werkzeugeinsatz zu wissenschaftlich belastbaren Ergebnissen führt, muss sich der Benutzer allerdings das Wissen um deren Funktionsweise aneignen, sonst kann er nicht beurteilen, ob die errechneten Ergebnisse statistisch valide sind.
6. Allerdings zeichnet sich auch das Szenario ab, dass die Benutzer digitaler Editionen irgendwann nicht mehr nur Menschen sind, sondern ‚lesende Maschinen‘, die vom Menschen nur noch in ihrer Textverarbeitung beobachtet werden (*Distant Reading*).⁸⁷
7. Parallel zu dieser völlig neuen Art des Lesens steht eine völlig neue Art des Edierens. Denkbar erscheint eine editorische Praxis, die, leistungsfähige OCR- (z. B. auch für frühe Drucke oder Handschriften) und KI-Modelle (z. B. für Schreiber- und Varianzanalysen, womöglich einst sogar für punktuelle Emendationen) vorausgesetzt,⁸⁸ weitgehend automatisiert abläuft.⁸⁹

⁸⁵ Robinson (Anm. 27), 5 f.

⁸⁶ Vgl. Ray Siemens, „Text Analysis and the Dynamic Edition? A Working Paper, Briefly Articulating Some Concerns with an Algorithmic Approach to the Electronic Scholarly Edition“, in: *CHWP* A.36 (2005), http://projects.chass.utoronto.ca/chwp/Casta02/Siemens_casta02.htm (letzter Aufruf 14.7.2017).

⁸⁷ Crane/Bamman/Jones (Anm. 67), 52.

⁸⁸ Zur Software-basierten Schreibererkennung vgl. diverse Artikel in Malte Rehbein/Patrick Sahle/Torsten Schaßan (Hg.), *Kodikologie und Paläographie im digitalen Zeitalter*, Norderstedt 2009: Mark Aussems/Axel Brink, „Digital Paleography“, 293–308; Maria Gurrado, „*Graphoskop*, uno strumento informatico per l’analisi paleografica quantitativa“, 251–259, und Wernfried Hofmeister/Andrea Hofmeister-Winter/Georg Thallinger, „Forschung am Rande des paläographischen Zweifels: Die EDV-basierte Erfassung individueller Schriftzüge im Projekt *DAmalS**“, 261–292, sowie im selben Sammelband zum Einsatz neuronaler Netzwerke zur automatisierten Transkription von Handschriften Daniele Fusi, „Aspects of Application of Neural Recognition to Digital Editions“, 175–195; zu lernenden Texten Crane/Bamman/Jones (Anm. 67), 39–46.

⁸⁹ Das Grazer Projekt *Fontes Civitatis Retisponensis* erstellt die Editionstexte bereits ‚halbautomatisch‘, vgl. Kropač (Anm. 66), 175.

Zum Teil mögen sich diese Szenarien einer Weiterentwicklung der digitalen Edition noch wie Science Fiction ausnehmen. Man wird sehen, wie die technische Entwicklung verläuft, aber auch, welche der Chancen, die sie bietet, die Philologen überhaupt ergreifen wollen. Schließlich tun die genannten Punkte nicht nur je für sich faszinierende neue Möglichkeiten auf, sondern sie führen im Einzelnen auch nicht selten – wie schon in Abschn. 1.3 angedeutet – auf Nutzungsweisen hin, die der Literaturwissenschaft bislang wesensfremd waren (vor allem Punkt 6 und 7): Weil die Literaturwissenschaft nicht nur eine Wissenschaft vom Lesen, sondern auch – wo immer es um (im weitesten Sinne) hermeneutische Fragen geht – eine lesende Wissenschaft ist, der das ‚konservative‘ Lesen Mittel und Zweck zugleich sein muss, mag es sein, dass sie, wenn sie auf digitale Editionen setzt, auf ein disziplinäres Paradoxon (oder eine disziplinäre Spaltung?) zusteuert. Im Umkehrschluss lässt sich mutmaßen, dass sich das ‚neue‘ Lesen, dessen letzte Konsequenz die Automatisierung des Lesens ist, auf dem Feld der Literaturwissenschaft nicht flächendeckend durchsetzen wird. Mutmaßen lässt sich darüber hinaus, dass wissenschaftssoziologisch begründete Widerstände gegen die Tendenzen zur Entmächtigung der Herausgeberrolle, zur Öffnung der Edition nach außen und zum wachsenden Stellenwert des (statistisch, algorithmisch) Genäherten (vor allem Punkt 2 bis 5) nicht leicht zu überwinden sein werden.

Wenn diese Mutmaßungen zuträfen, müsste man die digitale Edition nicht als einen Ersatz oder die Perfektion der analogen Buchedition begreifen, die endlich möglich macht, was man sich lange gewünscht hat. Die digitale Edition wäre vielmehr einfach ein weiteres, zusätzliches, seiner Natur nach eigengesetzliches Angebot, das manchen literaturwissenschaftlichen Anliegen besser dienen wird als ihr traditionelles gedrucktes Gegenstück, anderen aber nicht.

Literatur⁹⁰

- Aland, Kurt/Aland, Barbara, *Der Text des Neuen Testaments. Einführung in die wissenschaftlichen Ausgaben sowie in Theorie und Praxis der modernen Textkritik*, Stuttgart ²1989 [1982].
- Altenhöner, Reinhard, „Trau, schau, wem. Zur Authentizität und Langzeitverfügbarkeit digitaler Objekte“, in: Christiane Fritze/Franz Fischer/Patrick Sahle u. a. (Hg.), *Digitale Edition und Forschungsbibliothek. Beiträge der Fachtagung im Philosophicum der Universität Mainz am 13. und 14. Januar 2011*, Wiesbaden 2011, 153–170.
- Aristeas, *Der König und die Bibel*. Griechisch/Deutsch, übersetzt und hg. von Kai Brodersen, Stuttgart 2008 [ca. 2. Hälfte des 2. Jahrhunderts v. Chr.].
- Aussems, Mark/Brink, Axel, „Digital Paleography“, in: Malte Rehbein/Patrick Sahle/Torsten Schaßan (Hg.), *Kodikologie und Paläographie im digitalen Zeitalter*, Norderstedt 2009, 293–308.
- Bein, Thomas, *Textkritik. Eine Einführung in Grundlagen germanistisch-mediävistischer Editionswissenschaft. Lehrbuch mit Übungsteil*, Frankfurt a. M. u. a. ²2011 [2008].

⁹⁰ Sämtliche digitalen Referenzen wurden letztmalig am 19.8.2017 eingesehen.

- Bohnenkamp, Anne/Henke, Silke/Jannidis, Fotis u. a. (Hg.), *Johann Wolfgang Goethe. Faust. Historisch-kritische Edition*, Beta-Version 2, Frankfurt a. M./Weimar/Würzburg 2016, <http://beta.faustedition.net/>.
- Boot, Peter/Zundert, Joris van, „The Digital Edition 2.0 and The Digital Library. Services, not Resources“, in: Christiane Fritze/Franz Fischer/Patrick Sahle u. a. (Hg.), *Digitale Edition und Forschungsbibliothek. Beiträge der Fachtagung im Philosophicum der Universität Mainz am 13. und 14. Januar 2011*, Wiesbaden 2011, 141–152.
- Braun, Manuel/Glauch, Sonja/Kragl, Florian (Hg.), *Lyrik des deutschen Mittelalters. Digitale Edition*, <http://www.ldm-digital.de/>.
- Bryant, John, *The Fluid Text. A Theory of Revision and Editing for Book and Screen*, Ann Arbor 2002, 149–151.
- Cerquiglini, Bernard, *Éloge de la variante. Histoire critique de la philologie*, Paris 1989.
- Ciula, Arianna/Stella, Francesco (Hg.), *Digital Philology and Medieval Texts*, Ospedaletto 2007.
- Crane, Gregory/Bamman, David/Jones, Alison, „ePhilology. When the Books Talk to Their Readers“, in: Ray Siemens/Susan Schreibman (Hg.), *A Companion to Digital Literary Studies*, Malden, MA/Oxford/Carlton, Victoria 2007, 29–64.
- Czmiel, Alexander, *Sustainable Publishing. Standardization Possibilities for Digital Scholarly Edition Technology*, <http://dixit.uni-koeln.de/convention-2-abstracts/#czmiel>.
- Dahlström, Mats, „Digital Incunables. Versionality and Versatility in Digital Scholarly Editions“, in: Peter Linde/John W. T. Smith/Elena Emelianova (Hg.), *Electronic Publishing in the Third Millennium. Proceedings of an ICC/IFIP Conference Held at the Kaliningrad State University, Kaliningrad/Svetlogorsk, Russia, 17–19 August, 2000*, Washington 2000, 224–234.
- Dahlström, Mats, „Drowning by Versions“, in: *Human IT* 4/4 (2000), <https://humanit.hb.se/article/view/174/187>.
- Dahlström, Mats, „Editing Libraries“, in: Christiane Fritze/Franz Fischer/Patrick Sahle u. a. (Hg.), *Digitale Edition und Forschungsbibliothek. Beiträge der Fachtagung im Philosophicum der Universität Mainz am 13. und 14. Januar 2011*, Wiesbaden 2011, 91–106.
- Deegan, Marilyn/Sutherland, Kathryn u. a. (Hg.), *Jane Austen's Fiction Manuscripts*, <http://www.janeausten.ac.uk/index.html>.
- Div. Autoren, Kapitel „The New Philology“, in: *Speculum* 65/1 (1990), 1–108.
- Driscoll, Matthew J./Pierazzo, Elena (Hg.), *Digital Scholarly Editing. Theories and Practices*, Cambridge, UK, 2016, DOI <https://doi.org/10.11647/OBP.0095>.
- Finneran, Richard J. (Hg.), *The Literary Text in the Digital Age*, Ann Arbor 1996.
- Fredell, Joel, „Digital Philology in Little Boxes. Mobile Devices and The Book of Margery Kempe“, in: *Florilegium* 32 (2015), 77–98.
- Fusi, Daniele, „Aspects of Application of Neural Recognition to Digital Editions“, in: Malte Rehbein/Patrick Sahle/Torsten Schaßan (Hg.), *Kodikologie und Paläographie im digitalen Zeitalter*, Norderstedt 2009, 175–195.
- Gabler, Hans W., „Theorizing the Digital Scholarly Edition“, in: *Literature Compass* 7/2 (2010), 43–56.
- Gervais, Bertrand, „Is There a Text on This Screen? Reading in an Era of Hypertextuality“, in: Ray Siemens/Susan Schreibman (Hg.), *A Companion to Digital Literary Studies*, Malden, MA/Oxford/Carlton, Victoria 2007, 183–202.
- Giesecke, Michael, *Von den Mythen der Buchkultur zu den Visionen der Informationsgesellschaft. Trendforschungen zur kulturellen Medienökologie*, Frankfurt a. M. 2002.
- Girard, Danielle/Leclerc, Yvan, *Les Manuscrits de Madame Bovary. Edition intégrale sur le web*, <http://bovary.fr/>.
- Goethe, Johann W. von, *Goethes Werke*, hg. im Auftrage der Großherzogin Sophie von Sachsen. 144 Bde. Fotomechan. Nachdruck [Weimar 1887–1919], München 1987 [WA].
- Gumbrecht, Hans U., „Philology and the Complex Present“, in: *Florilegium* 32 (2015), 273–281.
- Gurrado, Maria, „Graphoskop, uno strumento informatico per l'analisi paleografica quantitativa“, in: Malte Rehbein/Patrick Sahle/Torsten Schaßan (Hg.), *Kodikologie und Paläographie im digitalen Zeitalter*, Norderstedt 2009, 251–259.

- Haberl, Doris, „Die Hofbibliothek Karls des Großen als Kristallisationspunkt der Karolingischen Renaissance. Geschichte, Umfeld, Wirkungen“, in: *Perspektive Bibliothek* 3.1 (2014), 111–139.
- Hofmeister, Wernfried/Hofmeister-Winter, Andrea/Thallinger, Georg, „Forschung am Rande des paläographischen Zweifels: Die EDV-basierte Erfassung individueller Schriftzüge im Projekt DAmalS*“, in: Malte Rehbein/Patrick Sahle/Torsten Schaßan (Hg.), *Kodikologie und Paläographie im digitalen Zeitalter*, Norderstedt 2009, 261–292.
- Kaiser, Reinhold (Hg.), *Die Glossen Ekkeharts IV. im Codex Sangallensis 621. Edition (Beta-version)*, <http://orosius.monumenta.ch/> (letzter Aufruf 19.8.2017).
- Kiernan, Kevin (Hg.), *Electronic Beowulf*, <http://ebeowulf.uky.edu/#>.
- Kraft, Herbert, *Editionsphilologie* [1990], Frankfurt a. M. u. a. ²2001.
- Kropač, Ingo H., „Work in Progress. Vom Digitalisat zum edierten Text“, in: Matthias Thumser/Janusz Tandecki (Hg.), *Editionswissenschaftliche Kolloquien 2005/2005. Methodik – Amtsbücher, Digitale Edition – Projekte*, Toruń 2008, 167–183.
- Kübler, Hans-Dieter, „Lesen und Medien in der zweiten Hälfte des 20. Jahrhunderts“, in: Rautenberg, Ursula/Schneider, Ute (Hg.), *Lesen. Ein interdisziplinäres Handbuch*, Berlin/Boston 2015, 793–812.
- Kuhn, Axel/Hagenhoff, Svenja, „Digitale Lesemedien“, in: Rautenberg, Ursula/Schneider, Ute (Hg.), *Lesen. Ein interdisziplinäres Handbuch*, Berlin/Boston 2015, 361–380.
- Kuhn, Axel/Kraus, Susanne, „Nutzergenerierte Texte in digitalen Netzwerken“, in: Rautenberg, Ursula/Schneider, Ute (Hg.), *Lesen. Ein interdisziplinäres Handbuch*, Berlin/Boston 2015, 679–699.
- Lavagnino, John, „Electronic Editions and the Needs of Readers“, in: W. Speed Hill (Hg.), *New Ways of Looking at Old Texts*, Bd. II: *Papers of the Renaissance English Text Society, 1992–1996*. Tempe, Arizona 1998, 149–156.
- Maas, Paul, *Textkritik*, Leipzig ³1960 [1927].
- Metzger, Bruce M., *The Text of the New Testament. Its Transmission, Corruption, and Restoration*, New York/Oxford ³1992 [1964].
- Mittler, Elmar/Rehbein, Malte, „Edition und Forschungsbibliothek. Chancen und Herausforderungen einer traditionsreichen Partnerschaft im digitalen Zeitalter“, in: Christiane Fritze/Franz Fischer/Patrick Sahle u. a. (Hg.), *Digitale Edition und Forschungsbibliothek. Beiträge der Fachtagung im Philosophicum der Universität Mainz am 13. und 14. Januar 2011*, Wiesbaden 2011, 9–21.
- Nichols, Stephen G., „Dynamic Reading of Medieval Manuscripts“, in: *Florilegium* 32 (2015), 19–57.
- O’Donnell, Daniel P., „Disciplinary Impact and Technological Obsolescence in Digital Medieval Studies“, in: Ray Siemens/Susan Schreibman (Hg.), *A Companion to Digital Literary Studies*, Malden, MA/Oxford/Carlton, Victoria 2007, 65–81.
- Perstling, Matthias P. (Bearb.), *Digitale Edition des steirisch-landesfürstlichen Marchfutterurbars von 1414/1426. Testversion*, <http://hfi.uni-graz.at/mfu/#fo674>.
- Petronius, *With an English Translation by Michael Heseltine. Seneca, Apocolocyntosis*, With an English Translation by W. H. D. Rouse, hg. von E. H. Warmington, Cambridge/London 1969.
- Pierazzo, Elena, „Modelling Digital Scholarly Editing. From Plato to Heraclitus“, in: Matthew J. Driscoll/Elena Pierazzo (Hg.), *Digital Scholarly Editing. Theories and Practices*, Cambridge 2016, 41–58, DOI <https://doi.org/10.11647/OBP.0095>.
- Price, Kenneth M., „Electronic Scholarly Editions“, in: Ray Siemens/Susan Schreibman (Hg.), *A Companion to Digital Literary Studies*, Malden, MA/Oxford/Carlton, Victoria 2007, 434–450.
- Rasmussen, Krista S. G., „Reading or Using a Digital Edition? Reader Roles in Scholarly Editions“, in: Matthew J. Driscoll/Elena Pierazzo (Hg.), *Digital Scholarly Editing. Theories and Practices*, Cambridge, UK, 2016, 119–133, DOI <https://doi.org/10.11647/OBP.0095>.
- Rautenberg, Ursula/Schneider, Ute (Hg.), *Lesen. Ein interdisziplinäres Handbuch*, Berlin/Boston 2015.

- Rehbein, Malte/Sahle, Patrick/Schaßan, Torsten (Hg.), *Kodikologie und Paläographie im digitalen Zeitalter*, Norderstedt 2009.
- Robinson, Peter M. W., „Is There a Text in These Variants?“, in: Richard J. Finneran (Hg.), *The Literary Text in the Digital Age*, Ann Arbor 1996, 99–115.
- Robinson, Peter, „Electronic Editions Which we Have Made and Which we Want to Make“, in: Arianna Ciula/Francesco Stella (Hg.), *Digital Philology and Medieval Texts*, Ospedaletto 2007, 1–12.
- Ross, Charles L., „The Electronic Text and the Death of the Critical Edition“, in: Richard J. Finneran (Hg.), *The Literary Text in the Digital Age*, Ann Arbor 1996, 225–231.
- Sahle, Patrick, „Die disziplinierte Edition – eine (kleine) Wissenschaftsgeschichte“, in: Matthias Thumser/Janusz Tandecki (Hg.), *Editionswissenschaftliche Kolloquien 2005/2005. Methodik – Amtsbücher, Digitale Edition – Projekte*, Toruń 2008, 35–52.
- Sahle, Patrick, „Digitale Editionstechniken und historische Quellen“, in: Stuart Jenks/Stephanie Marra (Hg.), *Internet-Handbuch Geschichte*, Köln/Weimar/Wien 2001, 153–166.
- Sahle, Patrick, „Digitale Editionstechniken“, in: Martin Gasteiner/Peter Haber (Hg.), *Digitale Arbeitstechniken für Geistes- und Kulturwissenschaften*, Wien/Köln/Weimar 2010, 231–249.
- Sahle, Patrick, „Digitales Archiv – Digitale Edition. Anmerkungen zur Begriffsklärung“, in: Michael Stolz/Lucas Marco Gisi/Jan Loop (Hg.), *Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien. Eine Standortbestimmung*, Zürich 2007, 64–84.
- Sahle, Patrick, „What is a Scholarly Digital Edition?“, in: Matthew J. Driscoll/Elena Pierazzo (Hg.), *Digital Scholarly Editing. Theories and Practices*, Cambridge, UK, 2016, 19–39, DOI <https://doi.org/10.11647/OBP.0095>.
- Sahle, Patrick, *A Catalog of Digital Scholarly Editions*. v. 3.0. Snapshot (2008 ff.), <http://www.digitale-edition.de/>.
- Sahle, Patrick, *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, 3 Bde., Norderstedt 2013.
- Schedl-Jokl, Maria, „Anmerkungen zur Geschichte der Weimarer Ausgabe“, in: Dies. (Hg.), *Goethes Werke*, Bd. 144: *Supplement zur Weimarer Ausgabe im Deutschen Taschenbuch-Verlag*, München 1987, 19–23.
- Shillingsburg, Peter L., *From Gutenberg to Google. Electronic Representations of Literary Texts*, Cambridge 2006.
- Siemens, R. G., „What Two Crowns Shall They Be? ‘Lower’ Criticism, ‘Higher’ Criticism, and the Impact of the Electronic Scholarly Edition“, in: W. Speed Hill (Hg.), *New Ways of Looking at Old Texts. III. Papers of the Renaissance English Text Society, 1997–2001*, Tempe, Arizona 2004, 37–46.
- Siemens, Ray, „Text Analysis and the Dynamic Edition? A Working Paper, Briefly Articulating Some Concerns with an Algorithmic Approach to the Electronic Scholarly Edition“, in: *CHWP A.36* (2005), http://projects.chass.utoronto.ca/chwp/Casta02/Siemens_casta02.htm.
- Stäcker, Thomas, „Creating the Knowledge Site. Elektronische Editionen als Aufgabe einer Forschungsbibliothek“, in: Christiane Fritze/Franz Fischer/Patrick Sahle u. a. (Hg.), *Digitale Edition und Forschungsbibliothek. Beiträge der Fachtagung im Philosophicum der Universität Mainz am 13. und 14. Januar 2011*, Wiesbaden 2011, 107–126.
- Stäcker, Thomas, „Konzepte zur Bereitstellung digitalisierter frühneuzeitlicher Quellen“, in: Rainer Hering u. a. (Hg.), *Forschung in der digitalen Welt. Sicherung, Erschließung und Aufbereitung von Wissensbeständen. Tagung des Staatsarchivs Hamburg und des Zentrums Geisteswissenschaften in der digitalen Welt an der Universität Hamburg am 10. und 11. April 2006*, Hamburg 2006, 143–152.
- Stella, Francesco, „Digital Philology, Medieval Texts, and the Corpus of Latin Rhythms, a Digital Edition of Music and Poems“, in: Ders./Arianna Ciula (Hg.), *Digital Philology and Medieval Texts*, Ospedaletto 2007, 223–249.
- Stock, Markus, „Introduction. Philological Moves“, in: *Florilegium* 32 (2015), 1–17.

- Stolz, Michael, „Benutzerführung in digitalen Editionen. Erfahrungen aus dem Parzival-Projekt“, in: Christiane Fritze/Franz Fischer/Patrick Sahle u. a. (Hg.), *Digitale Edition und Forschungsbibliothek. Beiträge der Fachtagung im Philosophicum der Universität Mainz am 13. und 14. Januar 2011*, Wiesbaden 2011, 49–80.
- Stolz, Michael, „Von den Fassungen zur Eintextedition. Eine neue Leseausgabe von Wolframs Parzival“, in: Dorothea Klein (Hg.), *Überlieferungsgeschichte transdisziplinär. Neue Perspektiven auf ein germanistisches Forschungsparadigma*, Wiesbaden 2016, 353–388.
- Taylor, Andrew, „Getting Technology and Not Getting Theory. The New Philology after Twenty-Five Years“, in: *Florilegium* 32 (2015), 131–155.
- Universitätsbibliothek Heidelberg/Universität Heidelberg SFB ‚Materiale Textkulturen‘ (Hg.), *Welscher Gast digital*, <http://digi.ub.uni-heidelberg.de/wgd/>.
- Vandendorpe, Christian, „Reading on Screen. The New Media Sphere“, in: Ray Siemens/Susan Schreibman (Hg.), *A Companion to Digital Literary Studies*, Malden, MA/Oxford/Carlton, Victoria 2007, 203–215.
- Vanhoutte, Edward, „Defining Electronic Editions. A Historical and Functional Perspective“, in: Willard McCarty (Hg.), *Text and Genre in Reconstruction. Effects of Digitalization on Ideas, Behaviours, Products and Institutions*, Cambridge, UK, 2010, 119–144.
- Vliet, H. T. M. van, „Die elektronische Edition als Lösung alter editorischer Probleme? Beobachtungen eines Editors“, in: Václav Bok/Frank Shaw (Hg.), *Magister et amicus. Fs. für Kurt Gärtner zum 65. Geburtstag*, o.O. 2003, 269–288.
- Zundert, Joris van, „Barely Beyond the Book?“, in: Driscoll/Pierazzo (Hg.), *Digital Scholarly Editing. Theories and Practices*, Cambridge, UK, 2016, 83–106, DOI <https://doi.org/10.11647/OBP.0095>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.



Modellierung von Textgeschichte. Bedingungen digitaler Analyse und Schlussfolgerungen für die Editorik

Gerrit Brüning

Editoren und Editionswissenschaftler haben immer wieder beklagt, die von ihnen vorgelegten Erkenntnisse und die Ergebnisse editionsphilologischer Arbeit stießen in der Literaturwissenschaft außerhalb der editorischen Teildisziplin kaum oder gar nicht auf Interesse.¹ Positiv gewendet: In kritischen Editionen und insbesondere ihren Apparaten liegen reiche, qualitativ hochwertige Daten über Texte und ihre Entwicklung vor; und obwohl die meisten dieser Daten kanonische Texte betreffen, da typischerweise nur kanonisierte Autoren mit einer kritischen oder gar historisch-kritischen Edition gewürdigt werden, harren die Daten größtenteils noch der literaturwissenschaftlichen Untersuchung. Woran liegt es, dass dieses Material weitgehend ungenutzt und damit im Grunde auch unbekannt geblieben ist?

1 Interpretation ohne Textkritik

Glaubt man dem editionswissenschaftlichen Diskurs, ist der Grund dafür in den Editionen selbst zu suchen: der für den Leser unzugänglichen technischen Form herkömmlicher (gedruckter) Apparate. Typisch für diese Sichtweise ist die Rede

¹Fredson Bowers, *Textual and literary criticism*, Cambridge, UK, 1959, 6; Hans Zeller, „Fünfzig Jahre neugermanistischer Edition. Zur Geschichte und künftigen Aufgaben der Textologie“, in: *Editio* 3 (1989), 1–17, hier: 15; Rüdiger Nutt-Kofoth, „Wie werden neugermanistische (historisch-)kritische Editionen für die literaturwissenschaftliche Interpretation genutzt? Versuch einer Annäherung aufgrund einer Auswertung germanistischer Periodika“, in: Thomas Bein (Hg.), *Vom Nutzen der Editionen. Zur Bedeutung moderner Editorik für die Erforschung von Literatur- und Kulturgeschichte*, Berlin 2015, 233–245, hier: 234 mit Hinweisen auf die ältere Diskussion.

G. Brüning (✉)

Wissenschaftlicher Mitarbeiter, Goethe- und Schiller-Archiv, Abteilung Medienbearbeitung und -nutzung, Klassik Stiftung Weimar, Weimar, Deutschland

E-Mail: Gerrit.Bruening@klassik-stiftung.de

vom sog. Variantenfriedhof. Sie impliziert, dass wissenschaftliche Apparate nur totes Wissen anhäufen, statt die Auseinandersetzung mit literarischen Texten zu befruchten. Vermutlich aber nicht aus diesem Grund meiden die meisten interpretatorisch arbeitenden Literaturwissenschaftler die kritischen Editionen und deren Apparate; denn ihnen fehlt ja gerade der Kontakt zum editionswissenschaftlichen Diskurs, durch den sie von der Kritik an den Variantenfriedhöfen hätten erfahren können.

Interpreten suchen nach Informationen, die zum Verständnis eines Textes beitragen. Es ist offensichtlich, dass z. B. Kommentare, aber nicht, dass auch Apparate – also Zeugenbeschreibungen und Variantenverzeichnisse usw. – dies tun. Man könnte nun versuchen, den auf den ersten Blick schwer erkennbaren Nutzen der editorischen für die interpretatorische Arbeit nachzuweisen (dazu später mehr). Stattdessen soll hier zunächst die für Interpreten typische Perspektive genauer beschrieben werden: Interpreten nehmen einen Text gewöhnlich als etwas unhintergebar Gegebenes hin. Zu textkritischen Problemen haben sie ein agnostisches Verhältnis: Sie tun gewissermaßen so, als könnte man über die Entstehung, Überlieferung, editorische Bearbeitung und somit auch die Fehlerhaftigkeit eines Textes nichts wissen. Von einem hermeneutischen Standpunkt aus gibt es an einer solchen Haltung textkritischen Problemen gegenüber zunächst nichts auszusetzen: Denn was sollte illegitim daran sein, den Text so zu interpretieren, wie er eben vorliegt, und ganz davon zu abstrahieren, warum er so ist, und ob er einmal anders war?² Wer nicht so verfährt, könnte man argumentieren, verwechselt im Grunde das Verstehen mit dem Erklären.³ Der Rückgriff auf frühere Autorvarianten als Interpretament kann sogar, wie man weiß, methodisch fragwürdig sein.⁴

Tatsächlich dürfen Interpreten den Text aber nicht ohne jede Rücksicht einfach so nehmen, wie er ist. Das zeigt sich, wenn es um Merkmale des Textes geht, die auf eine andere Instanz als den Autor zurückgehen.⁵ Man stelle sich vor, im Rahmen einer Interpretation des Kafka'schen *Processes* würde mit der Reihenfolge der Kapitel argumentiert: Eine solche Argumentation wäre dort von Grund auf verfehlt, wo die in den Editionen gewählte Kapitelreihenfolge selbst schon auf Interpretation beruht.⁶ Der Wortlaut einer Stelle mag einen denkbaren oder gar schönen Sinn ergeben: Sobald erwiesen ist, dass der Wortlaut durch den Fehler

²Vgl. Hans Zeller, „Textologie und Textanalyse. Zur Abgrenzung zweier Disziplinen und ihrem Verhältnis zueinander“, in: *Editio* 1 (1987), 145–158, hier: 148.

³Ebd., 148 f.

⁴Peter Szondi, „Über philologische Erkenntnis“, in: Ders., *Hölderlin-Studien. Mit einem Traktat über philologische Erkenntnis*, Frankfurt a. M. 1967, 9–34.

⁵Die folgenden Überlegungen gründen in Teilen auf internen Diskussionen zur Vorbereitung des konstituierten Texts des *Faust* der historisch-kritischen Faustedition.

⁶Franz Kafka, *Der Proceß*, hg. von Malcolm Pasley, in: *Schriften. Tagebücher. Kritische Ausgabe*, hg. von Jürgen Born/Gerhard Neumann/Malcom Pasley u. a., Frankfurt a. M. 1990, abgekürzt als *KKAP*, 2, 126 f.

eines Abschreibers oder Setzers zustande kam und auch nicht ex post vom Autor sanktioniert wurde, ist der sich daraus ergebende Sinn ignorabel. Ein Beispiel dafür ist *Faust I*, Vers 503, wo der Erdgeist von sich selbst sagt: „Wall’ ich auf und ab, / Webe hin und her!“⁷ In der zweiten und dritten Cotta’schen Ausgabe heißt es „Wehe“ statt „Webe“,⁸ was einen scheinbar ähnlich guten oder sogar überzeugenderen Sinn ergibt.⁹ Aufgrund des Lesartenbefunds ist jedoch anzunehmen, dass die Änderung auf den Setzer der zweiten Cotta’schen Ausgabe zurückgeht.¹⁰ Interpreten, die sich mit der Figur des Erdgeists beschäftigen und einen Text mit der Lesart „Wehe“ benutzen, werden diese Lesart für genauso autoritativ halten wie den umgebenden Wortlaut. Dass die Lesart „Wehe“ zwar ebenfalls historisch, aber nicht in demselben Sinn authentisch ist wie der unmittelbare Kontext und daher nicht mit der gleichen Berechtigung für die Charakteristik der Erdgeist-Figur herangezogen werden kann, bleibt ihnen verborgen. Und es wäre unangemessen, von ihnen zu verlangen, auf derartige Fehler nach dem sog. Überlieferungstechnischen Kriterium¹¹ gefasst zu sein.

Interpreten erwarten einen Text, den sie in seinen substanziellen Eigenschaften für authentisch, und das heißt aus ihrer Perspektive: für einen den erkenn- oder erschließbaren Intentionen des Autors möglichst nahe kommenden Text halten dürfen.¹² Dieser ist natürlich nicht identisch mit einem ‚textologisch‘ authentischen, d. h. überlieferungstreuen Text;¹³ von der Überlieferung haben die Interpreten, anders als vom Autor, vorläufig gar keinen Begriff. Anhand von Fehlern, die auf den Autor selbst zurückgehen, zeigt sich, dass Authentizität unter dem Gesichtspunkt der Interpretation sogar höher zu bewerten ist als die Freiheit von sachlichen Irrtümern. Zwei Beispiele aus Thomas Manns *Doktor Faustus* mögen das verdeutlichen: Serenus Zeitbloms Datierung des Beginns seiner Nieder-

⁷Anne Bohnenkamp/Silke Henke/Fotis Jannidis (Hg.), *Johann Wolfgang Goethe. Faust. Historisch-kritische Edition*, Version 1.1 RC, Frankfurt a. M./Weimar/Würzburg 2018, *Faust. Eine Tragödie*. Konstituierter Text. Bearbeitet von Gerrit Brüning/Dietmar Pravida, <http://v1-1.faustedition.net/> (letzter Aufruf 20.6.2019), hier: *Faust I*, V. 502 f.

⁸Ebd.

⁹Albrecht Schöne, *Johann Wolfgang Goethe. Faust. Kommentare*, Frankfurt a. M. 2017, 219.

¹⁰Gerrit Brüning, „Das Tageslicht. Zwei Berichtigungen einer Stelle in Goethes ‚Faust‘“, in: *Goethe-Jahrbuch* 130 (2013), 191–194, hier: 192.

¹¹Hans Zeller, „Struktur und Genese in der Editorik. Zur germanistischen und anglistischen Editionsforchung“, in: *LiLi* 19/20 (1975), 105–126, hier: 118.

¹²Peter Michelsen, „Diplomatik als Editionsprinzip. Zur Textgestalt des Faust anhand der Ausgabe Albrecht Schönes“, in: Ders., *Im Banne Fausts. Zwölf Studien*, Würzburg 2000, 192–206, hier: 193 f. Schon durch diese Erwartung an die Textgrundlage der Interpretation ist die Autor-Instanz für jede Interpretation elementar relevant, ohne dass zusätzlich noch auf der Ebene der Interpretation explizit oder implizit auf die Autor-Instanz bezuggenommen zu werden brauchte.

¹³Zeller (Anm. 2), 149.

schrift auf den 27. Mai 1943,¹⁴ die von Thomas Manns Tagebuch abweicht,¹⁵ und der musikwissenschaftlich unsinnige Ausdruck „Fugengewicht“ in Wendell Kretzschmars Vortrag über Beethovens Opus 111.¹⁶ Interpreten des *Doktor Faustus* erwarten an den betreffenden Stellen zu lesen, was der Autor des Romans, und nicht, was Mann in seinem Tagebuch und Adorno in seiner privaten Handreichung geschrieben hat – vorausgesetzt natürlich, dass es sich jeweils um mehr als bloße Schreibversehen, nämlich um Nachlässigkeiten oder Irrtümer des Autors, handelt.¹⁷

Selbstgewählte textkritische Ignoranz von Interpreten ist also nicht mit Indifferenz zu verwechseln. Interpreten setzen voraus, dass sie einen im beschriebenen Sinn authentischen Text interpretieren. Über interpretatorisch entscheidende Sachverhalte wie etwa den, dass die Kapitelreihenfolge des *Processes* in Teilen unsicher ist und auf editorischen Entscheidungen beruht, möchten sie an unübersehbarer Stelle informiert werden. Wissenschaftlich zitierfähige Ausgaben, d. h. auch brauchbare Lese- und Studienausgaben, zeichnen sich dadurch aus, dies zu leisten. Die Wahrscheinlichkeit, eine Interpretation durch Unkenntnis eines textkritischen Problems zu entwerten,¹⁸ ist daher gering, und die Risiken der textkritischen Ignoranz sind in der Praxis gut beherrschbar.¹⁹

Es ist Interpreten insofern nichts vorzuwerfen, wenn sie Apparate meiden. Die aus ihrer Sicht naheliegende Frage, was textkritische Erkenntnisse zum Verständnis eines Textes beitragen, geht jedoch in eine falsche Richtung, wenn damit zugleich nach dem Sinn und Zweck von Apparaten gefragt wird. Denn in einer solchen Frage würde sich ein limitierter Begriff von Literaturwissenschaft offenbaren, ein fehlendes Bewusstsein dafür, dass der Text selbst, sein Zustandekommen, seine Überlieferung und Fehlerhaftigkeit, von Interesse sein können – unabhängig davon, ob und was die auf diesem Gebiet gewonnenen Erkenntnisse zum Verständnis des Textes im klassisch interpretatorischen Sinn beitragen.

¹⁴Thomas Mann, *Große kommentierte Frankfurter Ausgabe. Werke – Briefe – Tagebücher*, hg. von Eckhard Heftrich u. a., Frankfurt a. M., abgekürzt als *GKFA*, 10.1: *Doktor Faustus. Das Leben des deutschen Tonsetzers Adrian Leverkühn, erzählt von einem Freunde*, hg. von Rupprecht Wimmer unter Mitarbeit von Stephan Stachorski, Frankfurt a. M. 2007, 11.

¹⁵*GKFA*, 10.2: *Kommentar*, hg. von Rupprecht Wimmer unter Mitarbeit von Stephan Stachorski, Frankfurt a. M. 2007, 174.

¹⁶*GKFA*, 10.2, 269.

¹⁷„Fugengewicht“ ist zwar auch Lese- und Abschreibfehler (*GKFA*, 10.2, 268 f.), aber unterlaufen konnte dieser nur infolge eines gleichzeitigen Sachirrtums (vgl. Hans Zeller, „Befund und Deutung. Interpretation und Dokumentation als Ziel und Methode der Edition“, in: Gunter Martens/Hans Zeller (Hg.), *Texte und Varianten*, München 1971, 45–89, hier: 62). Dass die Bearbeiter der *GKFA* dennoch „Eigengewicht“ in den Text gesetzt haben, bleibt aus dieser Perspektive zu bedauern.

¹⁸Zeller (Anm. 2), 149.

¹⁹Von den relativ wenigen Texten mit außergewöhnlich schwieriger Überlieferungslage abgesehen, ist mit einem erhöhten Risiko erst dann zu rechnen, wenn eine Interpretation auf Merkmale abhebt, die anfällig für autorfremde Einflüsse sind, z. B. den genauen Wortlaut weniger Einzelstellen, Schreibung und Interpunktion, Vers- oder Absatzenteilung, typographische Gestaltung etc.

24	A3 Ha Hb A4 Hc Hd A5	Es nur dunkel ◊ ◊ nur es Dunkel es nur sehen sehn	zu lernen , nicht auszuforschen ; Doch irrt ich ! doch
----	-------------------------------	--	---

Abb. 1 Variantenverzeichnis zum *Messias*²⁰

Wie stellt sich der von Bowers festgestellte ‚lack of contact‘²¹ zwischen ‚literary critics‘ und ‚textual critics‘ aus editionsphilologischer Perspektive dar? Erfreulicherweise hat man sich nicht darauf beschränkt, aufseiten der Interpreten ein schwach ausgeprägtes „Textbewußtsein“ zu beklagen oder ihnen falsche Interpretationen infolge mangelhafter Textgrundlagen und fehlender Variantenkenntnis vorzuhalten.²² Im Gegenteil: Die Entwicklung der Editionstechnik im 20. Jahrhundert lässt sich als ein immerwährendes Streben nach Lesbarkeit und Übersichtlichkeit trotz gleichzeitig größerer Genauigkeit und Vollständigkeit beschreiben. Die zuvor herrschende Trennung von Text und Apparat entfremdete die Varianten dem Kontext, in dem sie hätten verständlich werden können. Überwunden wurde diese Trennung durch das Prinzip der horizontal und zugleich vertikal oder syntagmatisch und paradigmatisch lesbaren Zeilensynopse (Abb. 1).²³

Der Nutzer blieb gleichwohl hypothetisch und die Adäquatheit im Verhältnis zum Gegenstand die eigentliche Richtschnur der Apparatdarstellung. Es gibt daher zwar keinerlei empirische Belege, wohl aber unabhängige Indizien dafür, dass mit der zeilensynoptischen Darstellung eine optimale Lesbarkeit erreicht wird: das bewährte Vorbild der musikalischen Partitur, mit der die synoptische Textdarstellung schon früh verglichen wurde,²⁴ und die bis heute andauernde Nutzung des TUSTEP-Programms *vergleich-aufbereite* in der editorischen Praxis. Vgl. das Beispiel in Abb. 2 aus der Arbeit an der Faustedition.²⁵

Auch längere Texte und eine größere Anzahl von Fassungen lassen sich auf diese Weise flüssig vergleichend lesen.²⁶ Synoptische Darstellungen haben jedoch

²⁰Weißraum bedeutet weitere Gültigkeit des darüberstehenden Textes, Rauten wie unter „Es nur“ zeigen Wegfall des darüberstehenden Textes an; Friedrich G. Klopstock, *Werke und Briefe. Historisch-kritische Ausgabe*, begründet von Adolf Beck/Karl L. Schneider/Hermann Tiemann, hg. unter der Leitung von Horst Gronemeyer, Berlin/New York 1974 ff., abgekürzt als *HKA*, hier: *Werke* IV 5.1, 3 (zu Gesang IV, V. 24).

²¹Bowers (Anm. 1), 6.

²²Vgl. die Zitate bei Nutt-Kofoth (Anm. 1), 235 f.

²³Vgl. Elisabeth Höpker-Herberg, „Überlegungen zum synoptischen Verfahren der Variantenverzeichnung. Mit einem Beispiel aus Klopstocks ‚Messias‘“, in: Martens/Zeller (Anm. 17), 219–232; Abb. 1.

²⁴Ebd., 219 f.

²⁵Bohnenkamp/Henke/Jannidis 2018 (Anm. 7).

²⁶Weitere Beispiele für den Einsatz des Programms *vergleich-aufbereite* zur Anzeige von Fassungsvarianten bietet die *Editorenwerkstatt Hesperus* (http://www.jean-paul-portal.uni-wuerzburg.de/neue_werkausgabe/editorenwerkstatt_hesperus/ [letzter Aufruf 6.5.2018]).

```

12.24 <stage+n="before_5199_d"> (derb, ungestüm und ungeschlacht) </stage>
[H.14] ===== (treten ein =====
[H.0a] ===== (treten ein =====

```

Abb. 2 Output des TUSTEP-Programms *vergleich-aufbereite* aufgrund eines Vergleichs von 2 I H¹⁴ und 2 I H^{0a} mit der hier als Grundtext gewählten Fassung 2 H²⁷

nichts daran ändern können, dass jegliche Form von Apparaten, geht man von Zeitschriften-Zitationen aus, allenfalls sporadisch benutzt wird.²⁸

2 Digitale Editionen: Nicht ohne Apparate

Dass reale Nutzer und ihr Umgang mit Editionen nicht greifbar sind, gilt zurzeit auch noch für digitale Editionen.²⁹ Deren Ziel ist, die den gedruckten Editionen attestierten Zugangshindernisse durch intuitive Usability zu überwinden und so einen größeren Nutzerkreis zu erreichen.³⁰ So wie im Fall gedruckter Editionen die Editoren und Rezensenten darüber befinden, welche Darstellungsform dem Anspruch der Übersichtlichkeit entspricht, liegt es bislang auch bei digitalen Editionen an den daran arbeitenden Philologen, Informatikern und Designern sowie einem kleinen Teil der interessierten Community, den Aufbau der Ansichten und die Navigationsstruktur im Hinblick auf optimale Benutzbarkeit zu beurteilen. Digitale Editionen ermöglichen es aber erstmals, genaue Kenntnisse über die tatsächliche Nutzung (besuchte Seiten, benutzte Links, Verweildauer, Downloads etc.) zu sammeln. Sollte es digitalen Editionen gelingen, editorische Erkenntnisse einer nicht-editorischen literaturwissenschaftlichen Perspektive zugänglich zu machen, würden digitale Editionen einen essenziellen Beitrag sowohl zur Legitimation editorischer Arbeit als auch zur Anschlussfähigkeit der *Digital Humanities* an die herkömmlichen Geisteswissenschaften leisten.

Ob dies gelingt, hängt von mehreren Bedingungsfaktoren ab. Zum einen natürlich von der weiteren Entwicklung der digitalen Editorik: Weder sind die Möglichkeiten digitaler Medien bislang ausgeschöpft, noch haben diese selbst ihre Entwicklung abgeschlossen. In digitalen Editionen ist daher auf Schritt und Tritt Darstellungsformen zu begegnen, die von denen der gedruckten Welt abgeleitet

²⁷ Gleichheitszeichen bedeuten Übereinstimmung mit dem Grundtext; vgl. Bohnenkamp/Henke/Jannidis 2018 (Anm. 7) und Gerrit Brüning, „Gültiger Wortlaut und ‚sinnliche Masse‘. Zur Textkonstitution des *Faust II*“, in: *Zeitschrift für deutsche Philologie* 138/3 (2018), 14, Anm. 56.

²⁸ Nutt-Kofoth (Anm. 1), XX24.

²⁹ Z. B. Bohnenkamp/Henke/Jannidis 2018 (Anm. 7).

³⁰ Anne Bohnenkamp/Silke Henke/Fotis Jannidis u. a., „Die digitale *Faust*-Edition. Zur neuen historisch-kritischen Ausgabe von Goethes Drama“, in: *Goethe-Jahrbuch* 133 (2016), 151–162, hier: 162.

	Ihr seht	ich	habe	nicht gesäumt		H6
	Ihr seht	ich	„hab mich	nicht gesäumt		H20
10555	Ihr seht	ich	„habe	nicht gesäumt		H20
	-	ich	habe	„freylich	nicht gesäumt	H20
	-	Ich	habe	freylich	nicht gesäumt	H13 H
	Die Waffen Säle ringsum		ausgeräumt			H6
	Die Waffensaäle ringsum		„aufgeräumt			H20
10556	Die Waffensaäle ringsum		„ausgeräumt			H20
	Die Waffensaäle ringsum		aufgeräumt	;		IV H ¹³ H
10557	Da standen sie zu Fuß		zu Pferde,			H ²¹
	Da standen sie zu Fuß		- zu Pferde,			H
10558	Als wären sie noch Herrn der Erde,					IV H ²¹ H

Abb. 3 Entwurf der zeugenübergreifenden Variantendarstellung für die Faustedition (2013)³²

sind, z. B. differenzierte Umschriften oder die erwähnten zeilensynoptischen Darstellungen (Abb. 3).³¹

Etwas grundsätzlich anderes zu erwarten, wäre momentan noch unangemessen: Digitale Editionen verstehen sich ja nicht als eine Ergänzung bereits bestehender oder gleichzeitig erarbeiteter gedruckter Editionen. Sie sind mit dem Anspruch verbunden, mindestens denselben Informationsgehalt abzubilden wie die gedruckten und für diese insofern einen vollwertigen Ersatz zu bieten. Es ist offensichtlich, dass digitale Editionen bereits jetzt mehr leisten und ermöglichen als gedruckte, z. B. bei der Verfügbarmachung umfangreicher Bildbestände, der Suche oder der Verknüpfung getrennt abgelegter Informationen. Es wäre aber ein Irrtum zu glauben, digitale Editionen würden gedruckte bereits in allen Teilen ersetzen können. Eine bleibende Herausforderung sind Darstellungsformen, die in der Lage wären, den Informationsgehalt der verschiedenen Apparatformen mit allen Komplexitäten der darin dargestellten Textentwicklung zu modellieren und zu visualisieren.

Man könnte einwenden, die Wiedergabe einzelner Fassungen im Rahmen einer digitalen Edition, die im Wesentlichen als Archiv konzipiert ist, mache den Apparat überflüssig; denn das, was der Apparat leiste – die Rekonstruierbarkeit einzelner Fassungen und darin vorgenommener Änderungen aus den im Apparat verzeichneten Varianten – erübrige sich durch die leichter nachvollziehbaren

³¹ Vgl. Bohnenkamp/Henke/Jannidis 2018 (Anm. 7) sowie Abb. 3.

³² Graue Schrift bedeutet weitere Gültigkeit des darüberstehenden Textes, Bindestriche wie unter „Ihr seht“ zeigen Wegfall des darüberstehenden Textes an. Zum Vergleich mit synoptischen Darstellungen in gedruckten Ausgaben siehe Abb. 1, zum Vergleich mit der für die Faustedition realisierten Umsetzung siehe <http://faustedition.net/print/faust.38#10555> (letzter Aufruf 20.5.2019) und die folgenden Verse. Vgl. auch die von Daniel J. Birnbaum entwickelte synoptische Edition der Nestorchronik (*Povest' vremennyx let*), <http://pvl.obdurodon.org/browser.xhtml#> (letzter Aufruf 6.5.2018).

Einzelwiedergaben. Dem ist jedoch aus zwei Gründen zu widersprechen: Erstens beschränkt sich die Leistung eines Apparats nicht darauf, einzelne Fassungen rekonstruierbar zu machen. Im Fall von Mehrfachüberlieferung aggregiert oder synthetisiert der Apparat die einzelnen Fassungen zu einer komplexen Information über die gesamte Textentwicklung. Er teilt zu jeder Stelle eines Textes mit, an welchen Stellen überhaupt Abweichungen zwischen Fassungen oder Änderungen in handschriftlichen Zeugen vorkommen und bringt diese in eine chronologische Reihenfolge, „damit so die allmähliche Entwicklung des Textes sich aufrolle“, wie Erich Schmidt zu seinem Apparat im *Faust II*-Band der Weimarer Ausgabe schreibt, die später oft unbedachterweise gerade dafür gescholten wurde, die Textentwicklung nicht darzustellen.³³ Erst wenn auch digitale Editionen diese Aggregation oder Synthetisierung leisten, können sie als vollwertiger Ersatz für die Apparate gedruckter Editionen gelten. Zweitens können bei der Wiedergabe einzelner Fassungen im Prinzip dieselben Komplexitäten auftreten wie bei der fassungsübergreifenden Textentwicklung. Hier stehen digitale Editionen daher prinzipiell vor denselben Problemen wie bei der Repräsentation fassungsübergreifender Varianten und zugleich unter einem erhöhten Druck, diese Probleme auch zu lösen. Denn nur was kodiert wird, kann auch visualisiert werden. Die Apparatformen der gedruckten Editionen sind zwar ebenfalls begrenzt, können sich im Bedarfsfall aber auf Behelfsmittel wie Zweitapparate zurückziehen.³⁴ Solche Anmerkungen mit verbalen Beschreibungen sind auch in digitalen Editionen möglich, widersprechen aber deren Konzept: Diese sollen Textveränderungen am Ende zwar möglichst ‚menschenslesbar‘ visualisieren, zunächst aber maschinenlesbar modellieren. Der Verzicht auf die Repräsentation von Apparatdaten würde also einen Informationsverlust mit sich bringen, ohne auf der anderen Seite den Modellierungsaufwand prinzipiell zu verringern.

Digitale Editionen stehen daher vor der Aufgabe, die in Apparaten kodierten Informationen medial transformiert zu repräsentieren. Damit ist natürlich nicht das typographische Erscheinungsbild herkömmlicher Apparate gemeint, sondern die darin typographisch kodierte Aussage über den Lesarten- oder handschriftlichen Befund. Doch ist dies überhaupt möglich? Sind alle in Apparaten enthaltenen Informationen einer maschinenlesbaren Modellierung zugänglich? Hier und da, z. B. in Reflexionen über den Unterschied zwischen einer typographischen gegenüber einer XML-basierten Kodierung von Informationen im digitalen Medium, heißt es, die Anforderung der Maschinenlesbarkeit zwingt zu einer größeren Eindeutigkeit und Formalisierung, als dies in gedruckten Editionen praktiziert worden sei. Für editionstechnisch anspruchsvolle typographische Darstellungen gilt dies jedoch, wenn überhaupt, am wenigsten. Man kann Schriftarten, Auszeichnungsformen, vertikal-horizontale Positionen, besondere Zeichen und dergleichen für

³³ *Goethes Werke*, hg. im Auftrage der Großherzogin Sophie von Sachsen, abgekürzt als *WA*, hier: I 15.2: *Lesarten zu Faust. Zweiter Theil*, bearbeitet von Erich Schmidt, Weimar 1888, 6.

³⁴ Vgl. Höpker-Herberg (Anm. 22), 232.

auf den ersten Blick weniger eindeutig und formalisiert halten als Bestandteile von (z. B.) XML-Markup. Dieser Eindruck wird sich jedoch zerstreuen, wenn man das den typographischen Darstellungen zugrundeliegende Regelwerk einbezieht. Umgekehrt sollte man sich durch die technische Anmutung von Markup-Sprachen nicht irreführen lassen: Markup bedarf einer Spezifikation und Interpretation, um die intendierten Bedeutungszuweisungen zu ermöglichen. Auch Unsicherheit und Vagheit entziehen sich nicht grundsätzlich den Ausdrucksmöglichkeiten von Markup-Sprachen (der entsprechende Ausdruck ist u. U. nur nicht so elegant, wie man das von der jeweiligen Markup-Sprache gewohnt ist). Es wäre verfehlt, das Problem der Modellierung von Apparatinformationen von Unsicherheit und Vagheit her zu denken, denn die allermeisten in Apparaten enthaltenen Informationen sind weder unsicher oder vage noch werden sie in uneindeutiger oder informeller Weise transportiert. Die echten Schwierigkeiten treten anderswo auf. Eine Angabe in einem Apparat mag so sicher, präzise und eindeutig wie nur irgendeine Angabe sein: Trotzdem (und dies betrifft nicht nur kuriose Rand- und Sonderfälle) kommt es vor, dass im gegenwärtig maßgeblichen Vokabular für die digitale Repräsentation solcher Informationen (TEI-XML) aus systematischen Gründen kein geeigneter Ausdruck zur Verfügung steht.³⁵

3 Informationsvermittlung und Relevanz

Angenommen, die betreffenden Probleme der Repräsentation von Apparaten wären sowohl auf der Ebene der Modellierung als auch auf der Ebene der Visualisierung gelöst, d. h. die medialen Möglichkeiten digitaler Editionen wären ausgeschöpft, die Vermittlung der in Apparaten enthaltenen Informationen wäre so intuitiv wie nur irgend möglich. Könnte dies allein schon die Gewähr dafür leisten, dass editorische Erkenntnisse einer nicht-editorischen literaturwissenschaftlichen (d. h. im eingangs beschriebenen Sinn interpretatorischen) Perspektive zugänglich werden? Dies anzunehmen hieße, zugunsten der fraglos wichtigen Informationsvermittlung völlig vom Charakter der Information, die da vermittelt wird, abzuweichen. Es muss daher gefragt werden, ob editorische Informationen von der Art sind, dass sich im üblichen Sinn interpretatorische Fragen an ihnen beantworten lassen. Sollten sie das nicht sein, liefe deren noch so intuitive Vermittlung ins Leere. Für typisch interpretatorische Fragestellungen sind textkritische Sachverhalte dann relevant, wenn sie die Deutung beeinflussen, wenn also unterschiedliche Lesarten an einer Stelle unterschiedliche Deutungen des Kontextes bis hin zum gesamten Text ermöglichen oder erzwingen. Stellt man sich Texte als System vor, das durch die Relationen seiner Komponenten konstituiert wird, und geht man davon aus, dass die Veränderung einer Komponente neue Relationen und

³⁵ Gerrit Brüning, „Putting variants first“, paper presented at *Connect, Animate, Innovate. TEI Conference and Members' Meeting 2015, Lyon*, 28.–31. October, Lyon 2015, Abstract: <http://tei2015.huma-num.fr/en/papers/> (letzter Aufruf 6.5.2018).

damit ein neues System entstehen lässt,³⁶ müsste jede Variante das Verständnis des Textes beeinflussen. Dieser Grundsatz ist jedoch zu pauschal, um auf konkrete interpretatorische Fragen fruchtbar angewendet werden zu können. Der Textbezug von Deutungen gestaltet sich nämlich nicht so, dass sie gleichmäßig auf alle Komponenten des Systems Bezug nehmen und infolgedessen von jeder Veränderung einzelner Komponenten berührt sind. Im Fokus stehen immer nur Teile des Textes, und je nach Granularität der Aussage kann vom genauen Wortlaut abstrahiert werden. In der Praxis schlagen unterschiedliche Lesarten nur ausnahmsweise auf die Deutung durch. Ein Beispiel dafür ist das Satzendezeichen in *Faust I*, V. 1685, demzufolge Fausts Satz

Doch hast du Speise die nicht sättigt, hast
[...]
Der Ehre schöne Götterlust,
Die, wie ein Meteor, verschwindet.³⁷

entweder (wie in der Überlieferung) als Aussagesatz oder (wie in älteren Editionen im Gefolge der *Weimarer Ausgabe*) als Fragesatz aufgefasst werden kann („[...] verschwindet?“³⁸). Davon wird die Deutung der Intentionen der Figur Faust im Wettgeschehen beeinflusst und davon wiederum der dramatische Konflikt insgesamt.

Es ließe sich eine Reihe weiterer Fälle anführen. Eine auf solchen Fällen beruhende Beweisführung, dass textkritische Probleme und editorische Entscheidungen generell für die Interpretation literarischer Texte relevant sind, bliebe jedoch anekdotisch. Selbst wenn es gelänge, auf diese Weise ein Interesse an dem in digitalen Editionen aufbereiteten Lesartenbefund zu wecken, so würden diese Erwartungen enttäuscht. Die Interpreten fänden die ‚Sonntagsvarianten‘ kaum unter den viel zahlreicheren unspektakulären ‚Werktagsvarianten‘. Deswegen ist zu vermuten, dass auch künftige optimale digitale historisch-kritische Editionen mit ihren Apparaten oder deren Äquivalenten zunächst auf dieselben Vermittlungsprobleme stoßen werden wie die gedruckten der Vergangenheit.

4 Vom Nutzen des Studiums der Varianten: Lessings 19. Literaturbrief

Es kommt deswegen darauf an, ein literaturwissenschaftliches Interesse an Apparatinhalt nicht von typisch interpretatorischen Fragestellungen, sondern vom editorischen Datenmaterial selbst her zu entwickeln, d. h. von der Gesamtheit der Informationen über die Entstehung und weitere Entwicklung von Texten.

³⁶Zeller (Anm. 11), 115.

³⁷Bohnenkamp (Anm. 7), *Faust I*, V. 1678–1685.

³⁸Dazu Gerrit Brüning, „Die Wette in Goethes Faust“, in: *Goethe Yearbook* 17 (2010), 31–54, hier: 41 f.

Das ist z. B. dadurch möglich, dass Varianten in derselben Art wie der Text als Interpretanda behandelt, also selbst zum Gegenstand interpretatorischer Bemühungen gemacht werden.³⁹ Mit einem solchen Verfahren lassen sich aber nur kleinere Textmengen, v. a. Gedichte, bewältigen. Bei umfangreicheren Texten mit Mehrfach- oder komplexer Überlieferung wäre die Anzahl der Varianten zu hoch, um auf diese Weise behandelt zu werden. Dass gerade auch eine zunächst unüberschaubare Vielzahl von Varianten wert ist, untersucht zu werden, wurde schon im 18. Jahrhundert erkannt.⁴⁰ Am interessantesten und ergiebigsten in dieser Hinsicht ist Lessings 19. Literaturbrief, der Beobachtungen über die Textgeschichte des Klopstock'schen *Messias* anstellt. Gegenstand der Erörterung ist zunächst eine Rezension Nicolais der 1756 in Halle erschienenen Ausgabe des Werks. Lessing zog aber die Ausgabe von 1751 zum Vergleich heran und macht die mit der neuen Ausgabe Einzug haltenden Änderungen, die der Rezensent gar nicht bemerkt hatte, zum eigentlichen Thema des Literaturbriefs. Er möchte zeigen, dass es gewinnbringend ist, sich intensiv mit diesen Änderungen zu beschäftigen:

Veränderungen und Verbesserungen aber, die ein Dichter, wie *Klopstock*, in seinen Werken macht, verdienen nicht allein angemerkt, sondern mit allem Fleiße studieret zu werden. Man studieret in ihnen die feinsten Regeln der Kunst; denn was die Meister der Kunst zu beobachten für gut befinden, das sind Regeln.⁴¹

Lessings Worte wurden wiederholt zitiert, wenn es darum ging, die philologische Erforschung von Textvarianten zu legitimieren.⁴² Dabei fielen die Deutungen der Sätze durchaus konträr aus. Das eigentlich Interessante an Lessings Ausführungen sind aber seine Beobachtungen über die Varianten, zu deren Studium er nicht nur programmatisch aufruft (wie etwa auch Goethe das im *Litterarischen Sanscülottismus* in Bezug auf Wieland tut), sondern erste Ansätze liefert. Im Einzelnen beobachtet Lessing metrische Änderungen, Änderungen der Wortfügung (innerhalb davon insbesondere Auflösung von Partizipialkonstruktionen), Änderungen des Ausdrucks, Erweiterungen, Hinzufügungen und Streichungen.⁴³

Zwei der von Lessing benannten Veränderungen seien hier zunächst herausgegriffen (bevor später auch die übrigen zur Sprache kommen): die Auflösung von Partizipialkonstruktionen und die Änderungen des Ausdrucks.

³⁹Z. B. Gerrit Brüning, *Ungleiche Gleichgesinnte. Die Beziehung zwischen Goethe und Schiller 1794–1798*, Göttingen 2015, 263–277, über die Varianten des Prologs zum *Wallenstein*.

⁴⁰Johann J. Bodmer/Johann J. Breitinger, „Vorrede der Herausgeber [zur Opitz-Ausgabe] (1745)“, in: Rüdiger Nutt-Kofoth (Hg.), *Dokumente zur Geschichte der neugermanistischen Edition*, Tübingen 2005, 1–6, hier: 4 f.

⁴¹Gotthold E. Lessing, *Werke und Briefe in zwölf Bänden*, hg. von Wilfried Barner/Conrad Wiedemann, Frankfurt a. M. 1985–2003, abgekürzt als *Lessing FA*, hier: 4, 508.

⁴²Vgl. Steffen Martus, *Werkpolitik. Zur Literaturgeschichte kritischer Kommunikation vom 17. bis ins 20. Jahrhundert. Mit Studien zu Klopstock, Tieck, Goethe und George*, Berlin/New York 2007, 368 f.

⁴³*Lessing FA*, 4, 508–511.

Für die Auflösung von Partizipialkonstruktionen führt Lessing zwei Beispiele an.

Aus dem ursprünglichen Wortlaut (Gesang II, V., 241 f.)

Daß er [...]

In der ersten von Gott ihm gegebenen Herrlichkeit glänzte⁴⁴

werde in der späteren Fassung

[...]

In der Herrlichkeit glänzte, die ihm der Donnerer anschuf. (ebd.)⁴⁵

Anstelle von (Gesang II, V. 281 f.)

„[...] Verkündigt der dampfende Nebel

Seine von allen Göttern so lange gewünschte Zurückkunft (ebd.)

liest Lessing in der späteren Fassung:

[...]

Seine Zurückkunft, auf welche die Götter so lange schon harren. (ebd.)⁴⁶

Die um die Partizipien „gegebenen“ und „gewünschte“ gebildeten Konstruktionen sind aufgelöst, und Lessing vermutet: Eben aus der Absicht, schwerfällige Partizipialkonstruktion nach Möglichkeit zu vermeiden, seien die Varianten zu erklären.⁴⁷ Da der umgebende Wortlaut ebenfalls von den Änderungen betroffen ist, sind allerdings auch andere Interpretationen möglich, Lessings Erklärung wurde angefochten.⁴⁸

Für die Änderungen des Ausdrucks nennt Lessing zwei Beispiele: Das Verb „wischen“ werde durch „trocknen“, das Substantiv „Behausung“ bis auf eine Ausnahme durch andere Ausdrücke ersetzt. In denselben Bereich fallen auch die durch religiöse Rücksichten erklärten Ausdrucksänderungen.⁴⁹

Die letztere Gruppe der lexikalischen Änderungen ist mit größerer Sicherheit zu beurteilen, wenn sie, wie in den beiden von Lessing angeführten Fällen von „wischen“,⁵⁰ nur den Bereich eines Worts umfassen, der umgebende Wortlaut also nicht betroffen ist. Die jeweilige Änderung hat aber dennoch einen Kontext:

⁴⁴ Lessing FA, 4, 509.

⁴⁵ Vgl. HKA, Werke, IV 4, 180 f.

⁴⁶ Vgl. ebd., 188 f.

⁴⁷ Vgl. auch Richard Hamel, *Klopstock-Studien*, Zweites Heft, Rostock 1880, 42 f.

⁴⁸ Eric A. Blackall, *Die Entwicklung des Deutschen zur Literatursprache 1700–1775. Mit einem Bericht über neue Forschungsergebnisse 1955–1964*, Stuttgart 1966 [engl. 1959], 436.

⁴⁹ Lessing FA, 4, 509.

⁵⁰ Ebd.

die analogen Änderungen an den übrigen Stellen, wo in der ursprünglichen Fassung „wischen“ oder „Behausung“ gestanden hatte. Dass letzteres Wort nicht nur an einigen Stellen, sondern mit einer Ausnahme überall ersetzt wurde, ist das entscheidende Indiz für die von Lessing festgestellte stellenübergreifende Bearbeitungstendenz. Genauso argumentiert Lessing im komplizierteren (und forschungsgeschichtlich strittigen) Fall der aufgelösten Partizipialkonstruktion. Lessing bricht die Serie der Beispiele mit den Worten ab: „Und so in hundert andern Stellen“.⁵¹ Seine Zuversicht bei der Beurteilung der beiden angeführten Varianten gründet also, anders als die Auffassung Blackalls,⁵² nicht auf einer differenzierten Betrachtung der Einzelstellen, sondern auf einer quantitativen Einschätzung des gesamten Lesartenbefunds.

5 Textentwicklung und Textgeschichte

Dies ist der Punkt, an dem die eigentlichen Stärken digitaler Editionen zum Tragen kommen, und zwar umso mehr, je umfangreicher die in ihnen enthaltenen Apparatdaten sind. Digitale Editionen schaffen die Voraussetzung dafür, dem im 18. Jahrhundert von Bodmer und Breitinger, Lessing und Goethe proklamierten Ziel des Studiums von Varianten näher zu kommen, als dies bislang möglich war. Einen wie großen Fortschritt dabei die im digitalen Medium erreichbaren intuitiven Präsentationsformen auch bedeuten: Sie allein sind es nicht, die den entscheidenden Beitrag dazu leisten. Ein noch wichtigerer Vorteil digitaler Editionen liegt darin, computergestützte Analysen der Text- und Apparatdaten zu ermöglichen, die dem menschlichen Leser bislang stumm blieben. In Kombination ergeben Text und Apparat die Gesamtheit der Veränderungen, die ein Text durchgemacht hat, soweit sie nachweisbar sind und für relevant erachtet werden. Die Gesamtheit dieser Veränderungen wird hier als ‚Textentwicklung‘ bezeichnet; der historische Prozess, an dem neben der Textentwicklung selbst auch die an der Textproduktion und -reproduktion beteiligten Akteure, materielle Gegebenheiten der Überlieferung, äußere Einflüsse, hypothetisch erschlossene Sachverhalte etc. Anteil haben, heißt ‚Textgeschichte‘. Der wesentliche Unterschied, durch den sich eine Analyse der Textgeschichte gegenüber der Wiedergabe der Textentwicklung als primärer Aufgabe des Apparats⁵³ auszeichnet, ist die Charakterisierung der Änderungen, die Suche nach Zusammenhängen zwischen ihnen und die Frage nach ihren Ursachen. Anhand von Lessings Argumentation ist das Zusammenspiel der verschiedenen Gesichtspunkte erkennbar. Er charakterisiert und gruppiert Änderungen, schließt von dieser Charakterisierung zunächst implizit auf den

⁵¹ Ebd.

⁵² Blackall (Anm. 48), 436.

⁵³ *Grundlagen der Goethe-Ausgabe* in Siegfried Scheibe, *Kleine Schriften zur Editionswissenschaft*, Berlin 1997, 256.

Akteur, dem sie zuzuschreiben sind (den Autor), und von dort aus explizit auf die Motive, von denen der Autor sich bei den Änderungen leiten ließ (teils Verbesserungsabsicht hinsichtlich des Metrums und Stils, teils idiosynkratische Neigungen, teils äußere Rücksicht auf religiöse Bedenken). Die computergestützte Analyse soll dazu dienen, auf Grundlage von Text- und Apparatdaten bestehende Hypothesen über die Geschichte von Texten überprüfbar zu machen, neue Hypothesen zu entwickeln und vergleichende Betrachtungen über Texte, Gattungen und Autoren hinweg zu ermöglichen.

„Textgeschichte“ ist von Haus aus eher ein Begriff der klassischen und altgermanistischen Philologie.⁵⁴ Dort bezieht sie sich üblicherweise auf eine Entwicklung des Textes, die zeitlich weit von dessen Entstehung entfernt und der Kontrolle des Autors entzogen ist und sich in späteren Abschriften niederschlägt, die meist nur indirekt von der ursprünglichen Fassung abstammen. Die Verhältnisse zwischen solchen Zeugen sind grundsätzlich von Abhängigkeit bestimmt, da jeder von ihnen eine mehr oder weniger treue Abschrift, d. h. eine Reproduktion von anderen ist. Neuere Überlieferungen stellen sich in vielfältiger Hinsicht anders dar. Hier sind wesentlich häufiger als in älteren Überlieferungen auch Entstehungshandschriften erhalten. Oft liegen mehrere Fassungen eines Textes vor, die nicht bloß Reproduktionen darstellen, sondern auf den Autor selbst zurückgehen und somit als prinzipiell gleichwertig anzusehen sind. Solche Fassungen sind vom jeweils früheren Zeugen nicht bloß abhängig, wie das für ältere Überlieferungen gilt. Denn eine Fassung kann zwar auf eine Vorlage zurückgehen und sie in diesem Sinn zumindest teilweise reproduzieren; meistens schlägt sich in ihr jedoch zugleich die verändernde Weiterarbeit des Autors am Text nieder.

Die neuere Überlieferung kennt Verhältnisse, die denen der älteren insofern analog sind, als in ihnen das reproduktive Moment vorherrscht: Das gilt insbesondere für die vom Autor nur punktuell kontrollierte Druckgeschichte. In Bezug auf die Druckgeschichte von Goethes Werken und im Einklang mit dem für die älteren Philologien typischen Gebrauch des Ausdrucks konnte Bernays 1866 formulieren, die „Geschichte des Textes“ sei bloß die „Geschichte seiner Corruptionen“.⁵⁵ Je stärker sich das Interesse der neugermanistischen Editionsphilologie auf die unmittelbar und ausschließlich auktorial bestimmte Schreibsituation fokussiert hat,⁵⁶ desto weniger Interesse kam der Textgeschichte im herkömmlichen, von den älteren Philologien hergeleiteten Sinn zu. Der Ausdruck

⁵⁴Kurt Gärtner, „Editorische Terminologie aus mediävistischer Sicht I“, in: Gunter Martens (Hg.), *Editorische Begrifflichkeit. Überlegungen zu einem „Wörterbuch der Editionsphilologie“*, Berlin/Boston 2013, 65–73, hier: 66; Winfried Woesler, „Althilologische editorische Terminologie in der Neuphilologie“, in: Ebd., 93–102, hier: 101; zum folgenden ähnlich Armin Hoenen/Gerrit Brüning, „Zur Stematologie neuerer Überlieferungen“, *DARIAH-DE Working Papers* 29, Göttingen 2019, URN urn:nbn:de:gbv:7-dariah-2019-1-3.

⁵⁵Michael Bernays, „Über Kritik und Geschichte des Goethe’schen Textes (1866)“, in: Nutt-Kofoth (Anm. 40), 16–26, hier: 23.

⁵⁶Klaus Hurlbusch, „Nachträgliche Bemerkungen zum Artikelentwurf „Textkritik (I)“ von Winfried Woesler“, in: Martens (Anm. 54), 230 f., hier: 230.

blieb in Gebrauch, änderte aber seine Bedeutung. Bei Bernays steht die aus Verderbnissen bestehende Textgeschichte den vom Autor herrührenden Änderungen („Geschichte des Werks“) gegenüber. Anstelle dieser qualitativen Unterscheidung findet man im zwanzigsten Jahrhundert eine zeitliche: Textgeschichte bezeichnet demnach die Veränderung, die ein Text nach dem Abschluss seiner Entstehungsgeschichte durchläuft.⁵⁷ Insgesamt scheint aber in neueren Verwendungen ein erweitertes Verständnis des Ausdrucks zu überwiegen: Als Textgeschichte wird die gesamte Entwicklungsgeschichte eines Textes bezeichnet, von seiner Entstehung an.⁵⁸ Stellenweise verschwimmt die Differenzierung zwischen Textgeschichte und Textentwicklung.⁵⁹

6 Textgeschichte und Textgenese

An die neueren Verwendungsweisen anschließend, wird unter ‚Textgeschichte‘ der historische Prozess der Textproduktion, Reproduktion und Tradierung in einem weiten Sinn verstanden. Dies schließt sowohl die Textgenese und damit im wesentlichen auktoriale Schreibvorgänge als auch die für die Textgeschichte – im herkömmlichen engeren Sinn – typischen Phänomene ein, d. h. die vom Autor nicht kontrollierten Veränderungen, die ein Text im Zuge seiner mechanischen Reproduktion durch Abschriften oder Drucke erfährt. Textgenese einerseits und Textgeschichte im herkömmlichen Sinn andererseits lassen sich ohnehin nur rein begrifflich sauber voneinander trennen, nicht in der realen Entstehungsgeschichte, da die für die herkömmliche Textgeschichte typischen Änderungen u. U. bis in die frühe handschriftliche Überlieferung zurückverfolgt werden können.⁶⁰ Textgenese und (herkömmliche) Textgeschichte interferieren so in mannigfacher Weise. Die Textgenese sollte der Textgeschichte im weiteren Sinn jedoch nicht einfach subsumiert werden. Ohne terminologisiert zu sein, führt der Ausdruck ‚Textgeschichte‘ einige Voraussetzungen mit sich, die seine Anwendung auf bestimmte Bereiche der Textgenese problematisch erscheinen lassen: Zunächst setzt der

⁵⁷ Gretchen L. Rogers, *Zur Textgeschichte der Gedichte Goethes*, Diss., Baltimore, MD, 1938; Siegfried Scheibe, „Von der Entstehungsgeschichte, der Textgeschichte und der zeitgenössischen Wirkungsgeschichte“, in: Siegfried Scheibe (Leitung)/Waltraud Hagen/Christel Laufer u. a. (Hg.), *Vom Umgang mit Editionen. Eine Einführung in Verfahrensweisen und Methoden der Textologie*, Berlin 1988, 160–204.

⁵⁸ Hans Zeller, „Zur gegenwärtigen Aufgabe der Editionstechnik. Ein Versuch, komplizierte Handschriften darzustellen (1959)“, in: Nutt-Kofoth (Anm. 40), 194–214, Woesler (Anm. 54), 101.

⁵⁹ Gunter Martens, „Textdynamik und Edition. Überlegungen zur Bedeutung und Darstellung variierender Textstufen“, in: Martens/Zeller (Anm. 17), 165–201.

⁶⁰ Anne Bohnenkamp, „Textkritik und Textedition“, in: Heinz Ludwig Arnold/Heinrich Detering (Hg.), *Grundzüge der Literaturwissenschaft*, München ³1999, 179–203, hier: 185; Zeller (Anm. 1), 12.

Begriff ‚Textgeschichte‘ die Existenz eines Textes voraus, dessen Entwicklung über materiell und zeitlich voneinander getrennte Stationen hinweg nachvollzogen werden kann. In einer frühen Phase der Genese ist aber der Text, der ja gerade erst entsteht, eventuell noch nicht greifbar. Es ist geradezu charakteristisch, dass der Anfangspunkt der Textgenese im Dunkeln liegt: „The beginning is almost never a fixed moment. [...] The search for these texts beginning leads to nowhere but an intertextual amalgam.“⁶¹ Textgenetische Aussagen beziehen sich denn auch oft auf Stadien, die kein materielles Korrelat hatten oder von denen keines überliefert ist (z. B. hypothetisch erschlossene Konzeptionen oder in Quellen bezeugte mentale Tätigkeiten). Von dort aus verfolgt die Untersuchung den Weg bis hin zu den Niederschriften, die „dichtungslogisch und strukturell bereits mögliche Bestandteile des entstehenden [...] Werks“ sind,⁶² obwohl sie mitunter noch nicht durch Textidentität auf spätere Fassungen bezogen werden können. Mit ihnen können Teile eines Werks durchaus schon als entstanden gelten, obwohl sie innerhalb der entstehungsgeschichtlichen Gesamtentwicklung bis zur ersten veröffentlichten Fassung ein noch frühes Stadium repräsentieren. Das Bindeglied zwischen der dunklen Vorgeschichte des Textes und solchen ersten Fassungen stellen Texte dar, die zu diesen letzteren in einem metatextuellen Verhältnis stehen. Solche Aufzeichnungen (oft als *Paralipomena* bezeichnet) lassen sich semantisch auf ein entstehendes Werk beziehen, stellen aber noch keine Entwurfsfassungen dar. Mit den ersten vorläufigen Entwurfsfassungen eines Textes beginnt das, was hier unter ‚Textgeschichte‘ verstanden wird. Sie erstreckt sich über die Folgentwürfe, Reinschriften, Typoskripte, Druckvorlagen, Korrekturbogen, Drucke, Errataverzeichnisse, Korrektorexemplare und weitere Drucke zunächst bis zu dem Punkt, an dem der Autor letztmalig Einfluss auf die Entwicklung seines Textes nimmt, und schließlich darüber hinaus bis hin zu postum entstandenen Stadien der Editionsgeschichte.

‚Textgenese‘ ist bei näherem Hinsehen also kein bloßer Unterbegriff von ‚Textgeschichte‘. Beide überlappen sich vielmehr: Die Textgenese reicht weiter zurück in Stadien, die zur Entstehungsgeschichte eines Werks gehören, die sich aber noch nicht als Text manifestieren, dessen Entwicklung über verschiedene Stationen hinweg nachvollzogen werden kann. Die Textgeschichte umgekehrt umfasst Stadien, die der Entstehung eines Textes zeitlich nachfolgen und daher nicht mehr im gebräuchlichen Sinn als genetisch gelten können – Stadien etwa, die aus der Weiterarbeit des Autors an einem bereits fertigen Text resultieren. Mit anderen Worten beginnt und endet die Textgenese jeweils früher als die Textgeschichte. Der bisherigen Verwendungsweise entsprechend wird auch hier auf eine Terminologisierung, wie sie andere Ausdrücke in der neugermanistischen Editionswissenschaft erfahren haben (‚Fassung‘, ‚Fehler‘ etc.), verzichtet.

⁶¹ Dirk van Hulle, *Textual Awareness. A Genetic Study of Late Manuscripts by Joyce, Proust, and Mann*, Ann Arbor 2004, 156.

⁶² Anne Bohnenkamp, „... das Hauptgeschäft nicht außer Augen lassend“. *Die Paralipomena zu Goethes Faust*, Frankfurt a. M. 1994, 828.

Der Ausdruck ‚Textgeschichte‘ ist am besten geeignet, Vorgänge wie die von Lessing beobachteten Änderungen am Text des *Messias* zu bezeichnen. Die drei Studien Richard Hamels von 1879/80 tragen denn auch den Titel *Zur Textgeschichte des Klopstock’schen Messias*.⁶³

Als Lessing seine Untersuchung der von ihm noch nicht so genannten Textgeschichte des *Messias* anstellte, tat er dies ohne eine wissenschaftliche Ausgabe, die ihm dazu die Grundlage bot. Er musste anhand von Druckexemplaren das Vorliegen von unterschiedlichen Fassungen und deren Varianten selbst ermitteln. Seine Überlegungen geben jedoch einen Eindruck davon, was sich mit den in der synoptischen Darstellung der Hamburger Klopstock-Ausgabe enthaltenen Informationen anfangen ließe, und auch, was daran aus einer allgemeineren literaturwissenschaftlichen Perspektive relevant sein könnte. Lessings Beobachtungen leiten sich nicht von einer interpretatorischen Fragestellung, sondern vom Lesartenbefund her, soweit er für ihn überschaubar war. Sie führen daher auf Merkmale, die vielleicht nicht in einer typischen Interpretation, wohl aber in einer Analyse der literarischen Form oder der Entwicklung der Literatursprache⁶⁴ zum Thema gemacht werden. In Teilen führen sie durchaus auch auf den Gehalt des Werks: sein Verhältnis zu den religiösen Tendenzen der Zeit, mit Konsequenzen für die Figurencharakteristik.⁶⁵ Wenn der Lesartenbefund sich dazu eignet, führt seine Untersuchung also ganz von selbst auf Erkenntnisse von allgemeiner literaturwissenschaftlicher Relevanz.

Jeder Text hat seine eigene Geschichte. Wenn ein Autor wie im Fall des *Messias* selbst eine Mehrzahl von Fassungen hervorbringt, sind diese von literaturgeschichtlichem Interesse, da Revisionen auf eine veränderte Konzeption des Werks oder einen poetologischen Wandel im Schaffen des Autors hindeuten können, aber auch auf Rezeptionsbedingungen und -formen reagieren. Nicht alle Autoren üben eine solche fortdauernde Werkherrschaft aus. Bereits Veröffentlichungen zu Lebzeiten können starken äußeren Einflüssen und Einschränkungen unterliegen. Für diese und umso mehr für erst postum veröffentlichte Werke gilt, dass zeitgenössische Herausgeber und Nachlassverwalter dem Werk ihren Stempel aufdrücken, indem sie verändernd in den Text eingreifen. Neuere wissenschaftliche Editoren lehnen solche autorfremden und zumal postum vorgenommene Veränderungen als inauthentisch ab, versuchen, diese Spuren der „sekundären Textgeschichte“⁶⁶ zu tilgen und so dem vom Autor hinterlassenen Text wieder möglichst nahezukommen. Entwicklungen dieser Art prägen die Rezeption von Texten und sind daher literaturwissenschaftlich ebenfalls von Interesse. So ist es

⁶³Vgl. Hamel (Anm. 47). Weil sie auch die gesamte Druckgeschichte reflektieren, würde es dem allgemein üblichen Verständnis von ‚Genese‘ widersprechen, die synoptische Apparatform der Hamburger Klopstock-Ausgabe oder ähnliche Formen für andere Autoren generell als ‚genetische Darstellungen‘ zu bezeichnen.

⁶⁴Z. B. Blackall (Anm. 48).

⁶⁵Lessing FA, 4, 511 f.

⁶⁶Zeller (Anm. 58).

etwa von Nutzen, die Geschichte der Brod'schen Kafka-Ausgabe und ihr Verhältnis zum handschriftlichen Nachlass auf der durch die aktuellen Editionen verbesserten Grundlage zu untersuchen.⁶⁷

Man darf allerdings nicht verlangen, der Lesartenbefund müsse Antworten auf beliebige literaturwissenschaftliche Fragen geben können. Die Fragen müssen vielmehr vom Lesartenbefund her entwickelt werden. Es sind dann aber nicht die editorischen Darstellungen, also Apparate, die auf diese Fragen Antwort geben. Sie sollen die Lesarten gerade nicht so sortieren, selegieren oder präsentieren, dass dem Nutzer eine bestimmte Perspektive oder Hypothese nahegelegt würde, sondern vorerst nur dokumentieren – unabhängig davon, welchen Gewinn der Nutzer daraus zu ziehen vermag.

7 Lessings Änderungstypen und ihre Operationalisierung

Der Nutzer sieht sich daher in einen „Urwald“ versetzt,⁶⁸ den er nur unter großen Anstrengungen durchqueren kann. Liegen die Informationen über die Textentwicklung aber in maschinenlesbarer Form vor, können digitale Methoden helfen, Schneisen durch das Dickicht zu schlagen. Dabei stellt sich dieselbe Frage wie bei computergestützten Textanalysen allgemein: Sind Programme in der Lage, Text- und Apparatdaten gewinnbringend zu interpretieren? Anders gefragt: Lassen sich Lessings Beobachtungen so operationalisieren, dass sie mit digitalen Analyseverfahren überprüft und diese Verfahren dann auf vergleichbare Daten angewendet und zur Bildung neuer Hypothesen genutzt werden können? Um dieser Frage nachzugehen, sei zunächst an die obige Aufzählung der von Lessing beobachteten Änderungen erinnert (Reihenfolge gemäß dem Text des 19. Literaturbriefs):

1. metrische Änderungen
2. Änderungen der Wortfügung, innerhalb davon insbesondere
 - a) Auflösung von Partizipialkonstruktionen
3. Änderungen des Ausdrucks
4. Erweiterungen
5. Hinzufügungen
6. Streichungen

Es versteht sich von selbst, dass alle Arten von Änderungen sich in einzelnen Fassungsunterschieden niederschlagen müssen, da die von Lessing beschriebenen Typen ja induktiv aus diesen gewonnen sind. Als Grundlage einer

⁶⁷Vgl. Eric L. Marson., „Die Prozeß-Ausgaben. Versuch eines textkritischen Vergleichs“, in: *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* 42 (1968), 760–772.

⁶⁸Schmidt (Anm. 33), 169.

computergestützten Analyse müssen nun zunächst die einzelnen Abweichungen ermittelt und maschinenlesbar kodiert werden.⁶⁹ Auf die ermittelten Abweichungen muss sodann ein Set von Regeln angewendet werden, anhand dessen eine Abweichung einer oder mehrerer oder auch keiner der genannten Gruppen zugeordnet wird. Die Regeln für die Gruppen 5 und 6 lassen sich am leichtesten formulieren: 5 bedeutet ‚Plustext‘, 6 eine Lücke in der späteren Fassung. Gruppe 4 umfasst Ersetzungsvarianten, bei denen der ersetzende Text umfangreicher ist als der ersetzte.

Die Mitglieder der Gruppe 3 sind schwieriger zu identifizieren. Soweit es sich um Einzelwortvarianten handelt, könnte versucht werden, diejenigen herauszufiltern, die sich nicht durch Interpunktion, Schreibung oder infolge einer grammatischen Änderung unterscheiden, bei denen die betreffenden Wortformen also auf unterschiedliche Lemmata zurückzuführen sind. Zwei Abweichungen gehören allerdings nicht ohne Weiteres in die von Lessing ausgemachte Gruppe, wenn einmal ein Wort mit dem Lemma ‚a‘ durch ein Wort mit dem Lemma ‚b‘, das andere Mal aber ein Wort mit dem Lemma ‚b‘ durch ein Wort mit dem Lemma ‚a‘ ersetzt wird. In diesem Fall muss der Gesamttext einschließlich seiner invarianten Teile im Hinblick auf Worthäufigkeiten untersucht werden, um festzustellen, ob die Häufigkeit der Vertretung einzelner Lemmata, wie Lessing das für „wischen“ oder „Behausung“ feststellt, signifikant abnimmt oder einzelne Lemmata ganz verschwinden. Benötigt wird dafür zunächst eine Lemmatisierung beider Fassungen und dann, auf Basis dieser Lemmatisierung, zwei Wortindizes mit Häufigkeitsangaben, die miteinander verglichen werden können.

Die Mitglieder der Gruppe 2 zu ermitteln ist ebenfalls nicht trivial. Dies gilt auch für die Untergruppe 2a. Als Partizipialkonstruktionen umfassen sie mehrere Wörter, ersetzt werden also nicht nur Partizipien („gegebenen“, „gewünschte“), und diese zugleich auch durch Verbformen, die auf andere Lemmata zurückzuführen sind („anschuf“, „harrten“). Als einzig unmittelbar greifbares Merkmal bietet sich der Wegfall von Partizipien entweder durch Tilgung oder durch Ersetzung mit einer Phrase ohne Partizipien an. Um Varianten dieser Art zu identifizieren, reicht ein Vergleich des Wortlauts nicht aus, es müssen grammatikalische Informationen einbezogen werden. Die Information über die Wortart führt aber noch nicht weiter, denn in den Fällen, die Lessing im Auge hat, werden die Partizipien adjektivisch gebraucht („von Gott ihm gegebenen“, „von allen Göttern so lange gewünschte“). Geht man von Wortarten (*Part of Speech*) aus, müsste zugleich auf die Information zugegriffen werden können, dass die betreffenden Adjektive von Verben abgeleitet sind. Bei allen von Lessing als Verbesserungen der Wortfügung interpretierten Varianten (von denen die Auflösung

⁶⁹Dies kann entweder durch automatische Kollation der beiden von Lessing benutzten Fassungen oder durch die Retrodigitalisierung des Apparats der HKA geschehen, in der die komplexe typographische Struktur der synoptischen Darstellung in eine maschinell auswertbare Kodierung übersetzt wird. Im Fall einer Retrodigitalisierung müssen dann die Unterschiede zwischen den beiden von Lessing benutzten Fassungen aus der Gesamtmenge der Varianten isoliert werden.

von Partizipialkonstruktionen eine Untergruppe darstellt) ist aber immerhin zu erwarten, dass sich die geänderte Syntax in geänderten Wortarten niederschlägt. Anstelle des Wortlauts könnte also die Information über die Wortart als Heuristik dienen.

Syntaktische Änderungen können allerdings unterschiedlichste Ursachen haben und auch als Nebeneffekte anders als syntaktisch motivierter Änderungen auftreten. Dazu gehören die von Lessing erwähnten metrischen Verbesserungen (Gruppe 1). Erkannt werden könnten diese an folgenden Merkmalen, durch die sich die spätere von der früheren Fassung unterscheidet:

- veränderte Anzahl der Silben im Vers,
- veränderte Verteilung von Hebungen und Senkungen,
- Abbau der Spannung zwischen natürlichem Wort-/Satz- und metrischem Akzent (Häufigkeit ‚metrischer Komplexität‘ im Sinne Klemens Bobenhausens),⁷⁰
- Realisierung von hexametrischen Zäsuren, d. h. anders gesetzte Wortgrenzen im Verhältnis zur metrischen Position.

Die ersten beiden der genannten Phänomene sind wieder auch als Nebeneffekt anders motivierter Änderungen denkbar und daher nur ein schwacher Anhaltspunkt. Grundsätzlich ist es auch wie schon im Fall der in Gruppe 2 fallenden Änderungen erforderlich, die Texte auf der Ebene nicht des Wortlauts, sondern einer geeigneten (hier: metrischen) Annotation zu vergleichen, um die der bloße Text erst angereichert werden muss. Die zielgerichtete Operationalisierung wird hier und teils auch an anderen Stellen dadurch erschwert, dass Lessing Bearbeitungstendenzen nur vage oder exemplarisch beschreibt. Zur Formulierung operationalisierbarer Aussagen müssen dann spätere Forschungen wie diejenigen Hamels herangezogen werden. Grundsätzlich ist es aber möglich, textgeschichtliche Hypothesen wie diejenigen Lessings zumindest teilweise so zu operationalisieren, dass sie mit digitalen Analyseverfahren überprüft werden können.

8 Textgeschichte: Modell, Analyse, Darstellung

Die angestellten Überlegungen entspringen einem Projekt, das aus drei Teilen besteht:

1. Modell. Dieser Teil enthält allgemeine Überlegungen darüber, wie textgeschichtliche Vorgänge begrifflich gefasst und digital repräsentiert und mit welchen Kategorien sie analytisch beschrieben werden können.
2. Analyse. Dieser Teil besteht aus Fallstudien auf der Grundlage von vorliegenden editorischen Daten.
3. Darstellung. Dieser Teil befasst sich mit der Frage, wie textgeschichtliche Vorgänge dargestellt werden können.

⁷⁰Vgl. <https://metricalizer.de/de/about/> (letzter Aufruf 6.5.2018).

In den Bereichen Analyse (2) und Darstellung (3) liegen Ergebnisse vor, die am ersten und dritten Akt des *Faust II* gewonnen wurden.⁷¹ Sie basieren im Wesentlichen auf automatischer Kollation und in Teilen auch auf computergestützter Analyse des Kollationsergebnisses.⁷² Die dort und noch in anderem Zusammenhang entwickelten Formen graphenbasierter Darstellung textgeschichtlicher Verhältnisse diskutieren Hoenen/Brüning.⁷³

Um einer computergestützten Analyse zugänglich zu sein, müssen editorische Informationen zuallererst digital vorliegen. Das sind zum einen digitale Editionen im engeren Sinn, d. h. solche, deren Produktion und Präsentation voll und ganz im digitalen Medium stattfinden. Digitale Präsentationsformen sind aus dem genannten Grund aber kein Erfordernis. Deswegen kommen neben Daten aus Retrodigitalisierung auch solche infrage, die bei der Herstellung reiner Druckausgaben anfallen und professionell archiviert werden. Die Daten lassen sich zunächst hinsichtlich ihres Informationsgehalts einteilen in

- a) Textfassungen ohne Varianten („Lesetexte“),
- b) Textfassungen mit integrierten Apparateinträgen, in denen Varianten unterschiedlicher Zeugen verzeichnet werden,
- c) Transkriptionen einzelner Zeugen mit Binnenvarianten infolge handschriftlicher Änderungen,
- d) vom Text getrennte Apparateinträge mit Referenz auf den Lesetext (dazu gehören mitunter auch längere getilgte oder ersetzte Passagen, zu denen es selbst wiederum Apparateinträge gibt).
- e) Der Inhalt der Apparateinträge (a und d) kann die Struktur von Transkriptionen (c) haben.

TEI-XML-Daten eignen sich in jeder Hinsicht am besten, weil sie verhältnismäßig leicht mit Transformationsroutinen auf Basis von Standardtechnologien (namentlich XSLT) weiterverarbeitet werden können. In der Realität ist die technische Form vorliegender und künftig erwartbarer Daten jedoch denkbar heterogen. Sie umfasst neben TEI-XML auch proprietäre und bereits historisch gewordene Formate, SGML, HTML, XML-ähnliches non-XML, proprietäres XML. Es liegt nahe, diese Formate zunächst zu TEI-XML zusammenzuführen, um so die Weiterverarbeit- und Abfragbarkeit mit Standardtechnologien zu ermöglichen. Dieser Schritt ist mit erheblichem technischen Aufwand verbunden, aber keine rein technische Angelegenheit, weil er im Grunde auf etwas Ähnliches hinausläuft wie das Bestreben der ehemaligen Textologie zur Vereinheitlichung von Editionen. Außerdem ist das Vokabular der TEI nicht mit dem Anspruch entwickelt worden,

⁷¹ Brüning (Anm. 26).

⁷² Vgl. ebd., 26, Anm. 105.

⁷³ Hoenen/Brüning (Anm. 54).

die in der neugermanistischen Editionsphilologie entwickelten Apparatdarstellungen abzubilden.⁷⁴ Angesichts der Komplexität der kodierten Sachverhalte ist die Auszeichnungssprache XML auch nicht grundsätzlich alternativlos.⁷⁵ Momentan müssen die beiden einschlägigen Kapitel der *TEI-Guidelines*⁷⁶ zur Einzelzeugenwiedergabe (Kap. 11) und zur Wiedergabe zeugenübergreifender Varianten (Kap. 12) als Ausgangspunkt gelten.⁷⁷ Zur adäquaten Repräsentation von Textgeschichte gehört neben der Kodierung von Texten zudem die Modellierung von Überlieferungsverhältnissen, d. h. von Beziehungen zwischen (je nach Lage und Perspektive unterschiedlich gearteten) Objekten, wofür sich am ehesten Graphen eignen.⁷⁸

Vorausgesetzt, die erwähnten Daten liegen wohlgeformt und in hinreichend einheitlicher Form vor, ist der Weg für den an Lessing exemplifizierten Ansatz der Operationalisierung von textgeschichtlichen Hypothesen frei. Das von Lessing beschriebene Szenario basiert auf dem Vergleich mehrerer (hier zweier) Fassungen ohne Binnenvarianten. Dies ist insofern eine relativ günstige Ausgangslage, als die zwei Textstufen, die miteinander verglichen werden, zwei jeweils synchrone Zustände darstellen und die festgestellten Varianten somit einen synchronen Zusammenhang bilden. Bei autographen Entstehungshandschriften ist das nicht der Fall: Die sog. Grundschicht ist keine Textstufe, und der größte Teil der von ihr ausgehenden Entstehungsvarianten lässt sich nicht auf den fragmentarischen oder gar gesamten Text hochrechnen, da viele Änderungen bereits während der Niederschrift vorgenommen wurden und somit früher als nachfolgende Teile der Grundschicht zu datieren sind. Infolge dieser Asynchronität der Änderungen ist es weniger wahrscheinlich, dass über den gesamten Text verteilte handschriftliche Änderungen einen Zusammenhang bilden. Typisch sind stattdessen lokale Verbände, die durch syntaktische Zusammenhänge konstituiert werden. Der ‚Text vor allen Änderungen‘

⁷⁴Die Apparatdaten im Düsseldorfer Heine- und im Grabbe-Portal sind denn auch nicht gemäß ihrer Apparatstruktur, sondern nach ihrem typographischen Erscheinungsbild wiedergegeben worden.

⁷⁵Diskutiert werden zurzeit range- und graphenbasierte Textmodelle: LMNL, d. h. ‚Layered Markup and Annotation Language‘, siehe Wendell Piez, ‚Hierarchies within range space. From LMNL to OHCO‘, in: *Balisage. The Markup Conference. Proceedings* (2014), <http://www.balisage.net/Proceedings/vol13/html/Piez01/BalisageVol13-Piez01.html>, und TAG, d. h. ‚Text As Graph‘, siehe, Ronald H. Dekker, ‚It’s more than just overlap. Text As Graph‘, in: *Balisage. The Markup Conference. Proceedings* (2014), <https://www.balisage.net/Proceedings/vol19/html/Dekker01/BalisageVol19-Dekker01.html> (letzter Aufruf der Links 6.5.2018).

⁷⁶<https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html> (letzter Aufruf 6.5.2018).

⁷⁷In Bezug auf Kap. 12 wird bewusst nicht davon gesprochen, dass dieses generell für die Repräsentation von Apparateinträgen einschlägig ist. Je nach Überlieferungslage und Inhalt der Apparateinträge kann es adäquater sein, diese nicht mit dem Modul ‚critapp‘ (Kritischer Apparat), sondern mit dem Modul ‚transcr‘ (Transkription) abzubilden.

⁷⁸Vgl. Hoenen/Brüning (Anm. 54).

ist allermeist ein Konstrukt ohne realen genetischen Ort, weil es ihn mit Sicherheit nicht gegeben hat.⁷⁹ Er kann daher nicht in der gleichen Weise mit dem ‚Text nach allen Änderungen‘ verglichen werden wie die frühere Fassung des *Messias* mit der späteren.

Trotzdem kann es Gesichtspunkte geben, unter denen es sinnvoll ist, die Grundsicht eines Textes zu konstruieren. Diese bildet zwar aus dem genannten Grund keinen validen Textzustand, sie kann aber trotzdem auf Merkmale hin analysiert werden, die für die Charakterisierung der Textentwicklung relevant sind. Ein Beispiel hierfür ist die Beobachtung, dass Kafka bei der Arbeit am *Schloß*-Roman Abweichungen von der schriftsprachlichen Norm bei der Überarbeitung des Manuskripts tilgte und im späteren Verlauf der Niederschrift sprachliche Eigentümlichkeiten zu vermeiden versuchte.⁸⁰ Allgemeiner und unabhängig vom konkreten Fall formuliert, ist es grundsätzlich erstrebenswert, die Entwicklung des Wortmaterials auf mögliche Muster zu untersuchen. Eine noch basalere Frage lautet, welche Wortformen in der Geschichte eines Textes vorkommen. Um eine solche Frage beantworten zu können, muss die Gesamtheit der textgeschichtlich existenten Wortformen (*Tokens*) in den Daten repräsentiert sein.

9 Token-Fragmente und Pseudo-Tokens in digitalen Texten: Transkription oder Edition? (Konsequenzen)

Man könnte glauben, diese Menge der Tokens sei mit digitalen Volltexten trivial gegeben. Doch das ist nicht uneingeschränkt der Fall. Digitale wissenschaftliche Editionen streben bekanntlich ein hohes Maß an Genauigkeit an. Im Fall von Überlieferung mit handschriftlichen Revisionen wird dieses Ziel gewöhnlich ohne Rücksicht auf die Frage verfolgt, welche Tokens als textgeschichtlich real angesehen werden können. Dies lässt sich am Beispiel eines Apparateintrags aus der Kritischen Kafka-Ausgabe zeigen, der für die Verzeichnung von handschriftlichen Revisionen typisch ist:

zu(r) (s > r)ichtigen⁸¹

⁷⁹Vgl. auch Thomas Burch/Stefan Büdenbender/Kristina Fink u. a., „Text[ge]schichten. Herausforderungen textgenetischen Edierens bei Arthur Schnitzler“, in: Katharina Krüger/Elisabetta Mengaldo/Eckhard Schumacher (Hg.), *Textgenese und digitales Edieren. Wolfgang Koepfens „Jugend“ im Kontext der Editionsphilologie*, Berlin/Boston 2016, 87–105, hier: 90.

⁸⁰Franz Kafka, *Das Schloß*, hg. von Malcolm Pasley, in: *Schriften. Tagebücher. Kritische Ausgabe*, hg. von Jürgen Born/Gerhard Neumann/Malcolm Pasley u. a., Frankfurt a. M. 1982, abgekürzt als *KKAS*, 2, 79 f.

⁸¹*KKAP*, 2, 162.

Den Kontext bildet die Phrase „Sie werden alles zur richtigen Zeit erfahren.“⁸² Dem Apparateintrag zufolge wurde das „r“ nach „zu“ eingefügt und das folgende „s“ durch „r“ überschrieben. Für den menschlichen Leser ergibt sich daraus, dass „zu“ durch „zur“ ersetzt und der Wortansatz „s“ zugunsten des nachfolgend Niedergeschriebenen aufgegeben wurde. Stellt man sich die durch den Apparateintrag repräsentierte Information in digitaler (z. B. TEI-kodierter) Form vor –

```
zu<add>r</add>
<subst>
  <del>s</del>
  <add>r</add>
</subst>ichtigen
```

–, so kommen die Tokens „zu“, „zur“, „s“ und „richtigen“ darin nicht explizit vor, und das Pseudo-Token „*sichtigen“ würde nicht explizit ausgeschlossen (Asterisk * hier für ein Pseudo-Token). Je nach Einstellung des Programms wäre es denkbar, nur die Tokens vor allen Änderungen („zu“, „*sichtigen“) oder nach allen Änderungen („zur“, „richtigen“) oder auch alle zwischen kritischen Zeichen stehenden Token-Fragmente zu indexieren („zu“, „*r“, „s“, „*r“, „*ichtigen“). Für den menschlichen Leser ist klar, welche Wörter hier existieren, und nur auf den ersten Blick erscheint dieses Wissen interpretatorisch oder hypothetisch im Vergleich zur Wiedergabe des handschriftlichen Befunds. Denn die an derselben Stelle aufeinander geschriebenen Zeichen können ja erst aufgrund des Wissens darüber, welche Wörter und Wortansätze an dieser Stelle sprachlich möglich sind, korrekt entziffert werden. Das Urteil über die inskribierten Wörter geht der Dokumentation des Befunds also voraus, anstatt erst im Anschluss an die und auf Basis der Dokumentation gebildet zu werden.

Der Apparat der Kritischen Kafka-Ausgabe hält die Informationen über die Tokens zurück, weil er, obwohl Apparat, sich als „Wiedergabe der handschriftlichen Befunde“ versteht⁸³ und diese nach Art einer möglichst genauen, d. h. zeichengetreuen Transkription verzeichnet. Verglichen mit diesem ‚Prinzip Transkription‘ würde das ‚Prinzip Edition‘ stattdessen darauf hinauslaufen, die Textentwicklung möglichst nachvollziehbar zu machen und als Variante eine einzige sofort getilgte Phrase, nämlich „zu s“ (offenbar für „zu seiner Zeit“) mitzuteilen.⁸⁴ Damit würde die Perspektive von sichtbaren inskriptionellen Niederschlägen auf die in diesem Fall betroffenen Einheiten sprachlicher Art (Wörter einschließlich Wortansätzen) verschoben und damit die sprachliche Seite des

⁸² *KKAP*, 1, 9.

⁸³ *KKAP*, 2, 10.

⁸⁴ Zu diesem Ansatz, Varianten zu verzeichnen, vgl. Gerrit Brüning/Katrin Henzel/Dietmar Pravida, „Multiple Encoding in Genetic Editions, The Case of Faust“, in: *Journal of the Text Encoding Initiative* 4 (2013), <https://doi.org/10.4000/jtei.697>.

Textbegriffs akzentuiert.⁸⁵ Ein Apparateintrag der Form „zu s >“ (> für sofortige Tilgung) würde weniger genau über den inskriptionellen Befund (Einfügung und Daraufschreibung statt, z. B., Streichung der Phrase „zu s“), aber dafür genauer über die in der Textentwicklung an dieser Stelle vorkommenden sprachlichen Einheiten (Tokens) informieren. Einer computergestützten Analyse editorischer Information würde die an sprachlichen Einheiten orientierte Form der Verzeichnung nach dem ‚Prinzip Edition‘ entgegenkommen. Denn die exemplarisch beschriebenen textgeschichtlichen Hypothesen beziehen sich typischerweise auf Wörter oder größere sprachliche Einheiten. Editorische Wiedergaben nach dem ‚Prinzip Transkription‘ bewegen sich dagegen über weite Strecken unterhalb der Token-Ebene und damit unterhalb des Levels, das für computergestützte Analysen üblicherweise relevant ist. Dies entspricht einem – nicht immer gut durchdachten – editorischen Genauigkeitsanspruch, restringiert aber zugleich automatisierbare Prozesse der Analyse (angefangen mit der Tokenisierung) und der davon abhängigen Anreicherung (z. B. Lemmatisierung, *POS Tagging*).

An gedruckten Editionen geht diese Kritik insofern vorbei, als diese nicht darauf ausgelegt sind, späteren computergestützten Analysen möglichst vorzuarbeiten. Vom editorischen Standpunkt aus mag dort der Verzicht auf Genauigkeit hinsichtlich der sprachlichen Einheiten eher zu verschmerzen sein als ein Verlust an Genauigkeit bei der Befundwiedergabe, zumal der Leser aus einer Wiedergabe nach dem ‚Prinzip Transkription‘ die sprachlichen Einheiten rekonstruieren kann, nicht aber umgekehrt aus einer Wiedergabe nach dem ‚Prinzip Edition‘ den detaillierten inskriptionellen Befund. Digitale Editionen verfolgen ein doppeltes Ziel: die genaue Wiedergabe des Befunds *und* die Eignung der Daten für computergestützte Analysen. Dass diese Ziele in Spannung treten können, ist bislang nicht reflektiert worden. Stattdessen herrscht in digitalen Editionen das ‚Prinzip Transkription‘ vor. Sie entsprechen so dem beschriebenen Genauigkeitsanspruch mit Blick auf die Wiedergabe des handschriftlichen Befunds, handeln sich aber zugleich die beschriebenen Schwierigkeiten für eine computergestützte Analyse ein und laufen so Gefahr, einen wesentlichen Vorteil digitaler gegenüber gedruckten Editionen teilweise gleich wieder zu verspielen. Die sich daraus ergebenden Schwierigkeiten betreffen nicht nur Untersuchungen im Anschluss an die Edition, sondern durchaus auch ein essenzielles, dabei für trivial und selbstverständlich gehaltenes Feature wie die Volltextsuche: Jede digitale Edition steht vor der Frage, wie einerseits alle textgeschichtlich realen Tokens gefunden werden können, andererseits aber keine Artefakte wie ungrammatische, nicht existente Tokens.⁸⁶ Um dies zu gewährleisten, ist bereits ein komplexer Analyseschritt

⁸⁵Vgl. Gerrit Brüning/Katrin Henzel/Dietmar Pravida, „On the dual nature of written texts and its implications for the encoding of genetic manuscripts“, in: Jan C. Meister (Hg.), *Digital Humanities 2012. Conference Abstracts*, Hamburg 2012, 131–134.

⁸⁶Vgl. das GitHub-Issue *Search solution for words with inline markup*, <https://github.com/faust-edition/faust-web/issues/196> (letzter Aufruf 6.5.2018).

notwendig, nämlich die korrekte Extraktion aller Tokens aus einer Kodierung, in der diese z. T. fragmentiert sind.

Es ist nicht zu erwarten, dass digitale Editionen sich vom ‚Prinzip Transkription‘ zugunsten des ‚Prinzips Edition‘ verabschieden werden.⁸⁷ Wie kann es gelingen, die Nachteile dieses Verfahrens aufzufangen, textgeschichtliche Vorgänge datentechnisch so zu modellieren, dass sie computergestützten Analysen optimal zugänglich sind? Zunächst muss eruiert werden, welche Probleme bei der Ermittlung der korrekten Tokens bei einer gegebenen Kodierung zu erwarten sind. Im obigen Beispiel aus der Kritischen Kafka-Ausgabe fehlt etwa die Kennzeichnung der Information „(s > r)“ als Sofortrevision. Diese Kennzeichnung wäre nötig, um den Token „s“ (Wortansatz) zu gewinnen und das Artefakt „*sichtigen“ auszuschließen. Wenn ein Wort an zwei Stellen geändert wurde, z. B. „(B)(w > ö)ses“,⁸⁸ fehlt u. U. eine Information darüber, wie das chronologische Verhältnis der Änderungen einzuschätzen ist. In diesem Fall ist klar, dass die Änderung „w > ö“ (ebenfalls eine Sofortrevision) und die Hinzufügung von „B“ als synchron behandelt werden müssen, damit die beiden Tokens „w“ (Wortansatz) und „Böses“ rekonstruiert werden können. In einer Minderheit von Fällen könnten Änderungen an zwei Stellen im Wort aber insgesamt drei Tokens ergeben. Letztlich sind also zusätzliche Annotationen erforderlich oder ein Set von heuristischen Regeln, nach der die entsprechenden Fallunterscheidungen automatisch gefällt werden können. Dafür wiederum ist notwendig, dass die Probleme, deren Lösung hier skizziert wird, auf philologischer Seite in ihrer Tragweite erkannt und durchdacht werden. Philologische Bearbeiter können sich nicht darauf zurückziehen, dass sie solche Entscheidungen und die zugrundeliegenden Interpretationen des Befunds vermeiden möchten. Denn bei der Weiterverarbeitung für eine funktionierende Suche oder für auf der Tokenisierung aufbauende analytische Vorhaben sind solche Entscheidungen unausweichlich, und sie werden nach Regeln gefällt. Editionsphilologen sollten an der Formulierung und Anwendung dieser Regeln mitwirken, das Digitale in der Digitalen Literaturwissenschaft also nicht an Informatiker oder Tools delegieren, sondern als wesentlichen Bestandteil der eigenen Arbeit begreifen.

Ergänzung 2021

Das skizzierte Vorhaben nahm 2018 durch konkrete Fallstudien praktisch Gestalt an. Der Schwerpunkt lag dabei auf Verfahren zur computergestützten Auswertung von Fassungsvergleichen.

Die im Beitrag ins Auge gefasste Untersuchung der Kafka-Ausgabe Max Brods (vgl. Abschn. 6) erwies sich als geeigneter Fall für eine regelbasierte Klassifikation von Varianten. Verwendet wurde der *Process*-Text der Kritischen Ausgabe

⁸⁷In der Faustedition wurde mit dem Konzept der ‚textuellen Transkription‘, vgl. Brüning/Henzel/Pravida (Anm. 84), der Weg in Richtung einer Wiedergabe nach dem ‚Prinzip Edition‘ zwar eingeschlagen, dies mündete aber nicht in eine durchgängige Kodierung auf Token-Ebene.

⁸⁸KKAP, 2, 174.

als Vertreter des Kafka'schen Manuskripts. Um den Text der Kritischen Ausgabe dem des Manuskripts soweit wie möglich anzunähern, wurden die in Apparatform verzeichneten Emendationen in Austauschabweisungen verwandelt und per Skript größtenteils zurückgenommen. Der so generierte Text wurde mit dem Text der Erstausgabe (1925, vertreten durch die Version im Deutschen Textarchiv) kollationiert. Zur regelbasierten Klassifikation diente ein an die Verhältnisse des *Process*-Textes angepasstes Skript, dessen ursprüngliche Form (mit gleichartiger Fragestellung entwickelt am Fall von Andy Weirs *The Martian*) mir Christoph Schöch zur Verfügung stellte.⁸⁹ Es erwies sich, dass Max Brods Eingriffe in den Text sich in der überwiegenden Zahl der Fälle durch das Bestreben erklären lassen, den Text des Manuskripts an die damals geltenden schriftsprachlichen Normen anzupassen (neben Interpunktion und Orthographie betrifft dies auch etwa Elisionen und österreichische Umlaute). Nur wenige Eingriffe berühren den Sinn des Textes. Somit verfuhr Brod mit dem Text insgesamt bei weitem schonender (konservativer), als in Anbetracht der geläufigen Urteile über Brods Praxis als Nachlassherausgeber zu erwarten gewesen wäre (Brod habe willkürlich in die Texte eingegriffen, Kafkas dichterische Sprache angetastet). Methodisch analog, aber auf anderer technischer Grundlage realisiert und optimiert ist eine Untersuchung von Johann Peter Eckermanns postum erfolgter Redaktion des vierten Teils von Goethes *Dichtung und Wahrheit*.⁹⁰

Ein zweites Verfahren der Auswertung von Kollationsergebnissen zielte darauf, das Verhältnis dreier Fassungen zueinander zu analysieren, von denen eine (A) jeweils den stemmatischen Vorläufer der beiden anderen (B, C) darstellt. Abweichungen vom Grundtext werden in die folgenden drei Gruppen eingeteilt:

1. B und C weichen gemeinschaftlich vom Grundtext A ab.
2. Nur B weicht an der betreffenden Stelle von A ab (C dagegen ist mit A identisch).
3. Nur C weicht von A ab (B ist mit A identisch).
4. B und C weichen auf unterschiedliche Weise von A ab.

Realisiert mit einem Skript von Wilhelm Ott, diente das Verfahren ursprünglich dazu, das Verhältnis des Erstdrucks von Goethes *Helena* in der Ausgabe letzter Hand (sog. Taschenausgabe, 1827) mit zwei Folgedruckten zu untersuchen: dem

⁸⁹Vgl. Erik Ketzan/Christof Schöch, "What Changed When Andy Weir's *The Martian* Got Edited?"; in: *Digital Humanities Conference Book of Abstracts*, Montréal 2017 sowie jetzt Erik Ketzan/Christof Schöch, "Classifying and Contextualizing Edits in Variants with Coletto. Three Versions of Andy Weir's *The Martian*", in: *Digital Humanities Quarterly* 15.4 (2021) mit Hinweisen auf weitere für diesen Beitrag relevante Literatur. Die Daten zu Kafka besorgte Sandra Krause. Thorsten Vitt half bei der Prozessierung der XML-Daten, Elisa Beshero-Bondar bei der Kollation mit CollateX.

⁹⁰Zu Verfahren und Ergebnissen vgl. Gerrit Brüning, „Eckermanns Redaktion des vierten Teils von *Dichtung und Wahrheit*. Annäherungen auf dem Weg traditioneller und digitaler Philologie“, in: Anne Bohnenkamp/Bernhard Fischer (Hg.), *Goethes Dichtung und Wahrheit. Beiträge zu Goethes autobiographischen Schriften*. Berlin, Boston 2022, 79–100, hier: 95–100.

Doppeldruck der Taschenausgabe sowie der Oktavausgabe letzter Hand (beide 1828). Ziel war, von gemeinschaftlichen Abweichungen der beiden Folgedrucke vom Erstdruck auf die verschollene Vorlage der beiden Folgedrucke zu schließen.⁹¹ Dieselbe Prozedur des Dreiervergleichs erwies sich als geeignet, um auch andere textgeschichtliche Verhältnisse zu analysieren: In der Editions-geschichte des *Processus* ließen sich damit zahlreiche Lesarten der zweiten *Process*-Ausgabe Max Brods (1935) als Rückkorrekturen nach der Handschrift erklären. Noch überraschender: An immerhin ca. 70 Stellen ist Brods Erstausgabe handschriftentreuer als der Text der Kritischen Ausgabe, was zum einen nahelegt, dass die Zugeständnisse der KKA an die Lesbarkeit mitunter weiter gehen als unbedingt notwendig, und zum anderen nochmals auf die konservativen Züge von Brods editorischer Praxis verweist. Der Dreiervergleich kann auch die Untersuchung von frühen Stadien der Textgeschichte erleichtern, indem er dabei hilft, in auktorial korrigierten Schreiberabschriften stehengebliebene Abschreibefehler zu identifizieren. Wenn A die Vorlage, B die Schreiberabschrift im unrevidierten Zustand der Grundschrift (abgeleitet aus der XML-Kodierung der handschriftlichen Änderungen) und C die letzte Stufe der Schreiberabschrift nach allen Revisionen ist (wieder abgeleitet aus der XML-Kodierung), dann unterliegen die Varianten der oben definierten ersten Gruppe (B und C weichen gemeinschaftlich vom Grundtext A ab) dem Verdacht, als unerkannte Abschreibefehler stehengeblieben und fortgeschleppt worden zu sein.⁹² Computergestützte Methoden geben so nicht nur der historischen Analyse, sondern damit auch der Textkritik und Textkonstitution neue und mächtige Mittel an die Hand, sodass die Analyse von Textgeschichte zugleich die Textgeschichte selbst weiterschreiben kann.

Literatur⁹³

- Bernays, Michael, „Über Kritik und Geschichte des Goethe’schen Textes (1866)“, in: Rüdiger Nutt-Kofoth (Hg.), *Dokumente zur Geschichte der neugermanistischen Edition*, Tübingen 2005, 16–26.
- Blackall, Eric A., *Die Entwicklung des Deutschen zur Literatursprache 1700–1775. Mit einem Bericht über neue Forschungsergebnisse 1955–1964*, Stuttgart 1966 [engl. 1959].
- Bodmer, Johann J./Breitinger, Johann J., „Vorrede der Herausgeber [zur Opitz-Ausgabe] (1745)“, in: Rüdiger Nutt-Kofoth (Hg.), *Dokumente zur Geschichte der neugermanistischen Edition*, Tübingen 2005, 1–6.

⁹¹ Vgl. Brüning (Anm. 26), 216, Anm. 105. Bei der Vorbereitung der Daten und der Aufbereitung der Ergebnisse half Hans-Werner Bartz.

⁹² Diesen und anderen Verfahren verdanken sich etliche Abweichungen des *Faust*-Texts von allen vorangegangenen textkritischen Editionen, vgl. Gerrit Brüning/Dietmar Pravida, „Editorischer Bericht“, in: Johann Wolfgang Goethe, *Faust. Eine Tragödie*. Konstituierter Text. Bearbeitet von Gerrit Brüning/Dietmar Pravida, Göttingen ²2018 (= Johann Wolfgang Goethe. *Faust*. Historisch-kritische Edition, Anm. 7), 487–571, hier: 552–554.

⁹³ Sämtliche digitalen Referenzen wurden letztmalig am 6.5.2018 eingesehen.

- Bohnenkamp, Anne/Henke, Silke/Jannidis, Fotis u.a. (Hg.), *Johann Wolfgang Goethe. Faust. Historisch-kritische Edition*, Version 1.1 RC, Frankfurt a.M./Weimar/Würzburg 2018, <http://v1-1.fauedition.net>.
- Bohnenkamp, Anne/Henke, Silke/Jannidis, Fotis u.a., „Die digitale *Faust*-Edition. Zur neuen historisch-kritischen Ausgabe von Goethes Drama“, in: *Goethe-Jahrbuch* 133 (2016), 151–162.
- Bohnenkamp, Anne, „... das Hauptgeschäft nicht außer Augen lassend“. *Die Paralipomena zu Goethes Faust*, Frankfurt a. M. 1994.
- Bohnenkamp, Anne, „Textkritik und Textedition“, in: Heinz Ludwig Arnold/Heinrich Detering (Hg.), *Grundzüge der Literaturwissenschaft*, München ³1999, 179–203.
- Bowers, Fredson, *Textual and literary criticism*, Cambridge, UK, 1959.
- Brüning, Gerrit, „Eckermanns Redaktion des vierten Teils von *Dichtung und Wahrheit*. Annäherungen auf dem Weg traditioneller und digitaler Philologie“, in: Anne Bohnenkamp/Bernhard Fischer (Hg.), *Goethes Dichtung und Wahrheit. Beiträge zu Goethes autobiographischen Schriften*. Berlin, Boston 2022, 79–100.
- Brüning, Gerrit, „Putting variants first“, paper presented at *Connect, Animate, Innovate. TEI Conference and Members' Meeting 2015, Lyon, 28.–31. October*, Lyon 2015, Abstract: <http://tei2015.huma-num.fr/en/papers/>.
- Brüning, Gerrit, *Ungleiche Gleichgesinnte. Die Beziehung zwischen Goethe und Schiller 1794–1798*, Göttingen 2015.
- Brüning, Gerrit, „Die Wette in Goethes *Faust*“, in: *Goethe Yearbook* 17 (2010), 31–54.
- Brüning, Gerrit, „Gültiger Wortlaut und ‚sinnliche Masse‘. Zur Textkonstitution des *Faust II*“, in: *Zeitschrift für deutsche Philologie* 138 (2018), 191–221.
- Brüning, Gerrit/Henzel, Katrin/Pravida, Dietmar, „Multiple Encoding in Genetic Editions. The Case of *Faust*“, in: *Journal of the Text Encoding Initiative* 4 (2013), <https://doi.org/10.4000/jtei.697>.
- Brüning, Gerrit, „Das Tageslicht. Zwei Berichtigungen einer Stelle in Goethes ‚*Faust*‘“, in: *Goethe-Jahrbuch* 130 (2013), 191–194.
- Brüning, Gerrit/Henzel, Katrin/Pravida, Dietmar, „On the dual nature of written texts and its implications for the encoding of genetic manuscripts“, in: Jan C. Meister (Hg.), *Digital Humanities 2012. Conference Abstracts*, Hamburg 2012, 131–134.
- Brüning, Gerrit/Dietmar Pravida, „Editorischer Bericht“, in: Johann Wolfgang Goethe, *Faust. Eine Tragödie*. Konstituierter Text. Bearbeitet von Gerrit Brüning/Dietmar Pravida, Göttingen ²2018 (= Johann Wolfgang Goethe. *Faust*. Historisch-kritische Edition, hg. von Anne Bohnenkamp/Silke Henke/ Fotis Jannidis), 487–571.
- Burch, Thomas/Büdenbender, Stefan/Fink, Kristina u.a., „Text[ge]schichten. Herausforderungen textgenetischen Edierens bei Arthur Schnitzler“, in: Katharina Krüger/Elisabetta Mengaldo/Eckhard Schumacher (Hg.), *Textgenese und digitales Edieren. Wolfgang Koepfens „Jugend“ im Kontext der Editionsphilologie*, Berlin/Boston 2016, 87–105.
- Dekker, Ronald H., „It's more than just overlap. Text As Graph“, in: *Balisage. The Markup Conference. Proceedings* (2014), <https://www.balisage.net/Proceedings/vol19/html/Dekker01/BalisageVol19-Dekker01.html>.
- Gärtner, Kurt, „Editorische Terminologie aus mediävistischer Sicht I“, in: Gunter Martens (Hg.), *Editorische Begrifflichkeit. Überlegungen zu einem „Wörterbuch der Editionsphilologie“*, Berlin/Boston 2013, 65–73.
- Goethe, Johann Wolfgang, *Goethes Werke*, hg. im Auftrage der Großherzogin Sophie von Sachsen, abgekürzt als WA, I 15.2: *Lesarten zu Faust. Zweiter Theil*, bearbeitet von Erich Schmidt, Weimar 1888.
- Hamel, Richard, *Klopstock-Studien*, Zweites Heft, Rostock 1880.
- Hoenen, Armin/Brüning, Gerrit, „Zur Stemmologie neuerer Überlieferungen“, *DARIAH-DE Working Papers* 29, Göttingen 2019, URN urn:nbn:de:gbv:7-dariah-2019-1-3.
- Höpker-Herberg, Elisabeth, „Überlegungen zum synoptischen Verfahren der Variantenverzeichnung. Mit einem Beispiel aus Klopstocks ‚*Messias*‘“, in: Gunter Martens/Hans Zeller (Hg.), *Texte und Varianten*, München 1971, 219–232.
- Hulle, Dirk van, *Textual Awareness. A Genetic Study of Late Manuscripts by Joyce, Proust, and Mann*, Ann Arbor 2004.

- Hurlebusch, Klaus, „Nachträgliche Bemerkungen zum Artikelentwurf „Textkritik (I)“ von Winfried Woessler“, in: Gunter Martens (Hg.), *Editorische Begrifflichkeit. Überlegungen zu einem „Wörterbuch der Editionsphilologie“*, Berlin/Boston 2013, 230 f.
- Kafka, Franz, *Schriften. Tagebücher. Kritische Ausgabe*, hg. von Jürgen Born/Gerhard Neumann/Malcom Pasley u.a., Frankfurt a. M. 1982–2013, abgekürzt als *KKA*.
- Kafka, Franz, *Der Proceß*, hg. von Malcolm Pasley, Frankfurt a. M. 1990, abgekürzt als *KKAP*.
- Kafka, Franz, *Das Schloß*, hg. von Malcolm Pasley, Frankfurt a. M. 1982, abgekürzt als *KKAS*.
- Ketzan, Erik/Schöch, Christof, „What Changed When Andy Weir’s *The Martian* Got Edited?“, in: *Digital Humanities Conference Book of Abstracts*, Montréal 2017.
- Ketzan, Erik/Schöch, Christof, „Classifying and Contextualizing Edits in Variants with Coletto. Three Versions of Andy Weir’s *The Martian*“, in: *Digital Humanities Quarterly* 15.4 (2021).
- Klopstock, Friedrich G., *Werke und Briefe. Historisch-kritische Ausgabe*, begründet von Adolf Beck/Karl L. Schneider/Hermann Tiemann, hg. unter der Leitung von Horst Gronemeyer, Berlin/New York 1974 ff., abgekürzt als *HKA*.
- Lessing, Gotthold E., *Werke und Briefe in zwölf Bänden*, hg. von Wilfried Barner/Conrad Wiedemann, Frankfurt a. M. 1985–2003, abgekürzt als *Lessing FA*.
- Mann, Thomas, *Große kommentierte Frankfurter Ausgabe. Werke – Briefe – Tagebücher*, hg. von Eckhard Heftrich u.a., Frankfurt a. M., abgekürzt als *GKFA*.
- Mann, Thomas, GKFA, Bd. 10.1: *Doktor Faustus. Das Leben des deutschen Tonsetzers Adrian Leverkühn, erzählt von einem Freunde*, hg. von Rupprecht Wimmer unter Mitarbeit von Stephan Stachorski, Frankfurt a. M. 2007.
- Mann, Thomas, GKFA, Bd. 10.2: *Kommentar*, hg. von Rupprecht Wimmer unter Mitarbeit von Stephan Stachorski, Frankfurt a. M. 2007.
- Marson, Eric L., „Die Prozeß-Ausgaben. Versuch eines textkritischen Vergleichs“, in: *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* 42 (1968), 760–772.
- Martens, Gunter, „Textdynamik und Edition. Überlegungen zur Bedeutung und Darstellung variierender Textstufen“, in: Gunter Martens/Hans Zeller (Hg.), *Texte und Varianten*, München 1971, 165–201.
- Martus, Steffen, *Werkpolitik. Zur Literaturgeschichte kritischer Kommunikation vom 17. bis ins 20. Jahrhundert. Mit Studien zu Klopstock, Tieck, Goethe und George*, Berlin/New York 2007.
- Michelsen, Peter, „Diplomatik als Editionsprinzip. Zur Textgestalt des Faust anhand der Ausgabe Albrecht Schönes“, in: Ders., *Im Banne Fausts. Zwölf Studien*, Würzburg 2000, 192–206.
- Nutt-Kofoth, Rüdiger, „Wie werden neugermanistische (historisch-)kritische Editionen für die literaturwissenschaftliche Interpretation genutzt? Versuch einer Annäherung aufgrund einer Auswertung germanistischer Periodika“, in: Thomas Bein (Hg.), *Vom Nutzen der Editionen. Zur Bedeutung moderner Editorik für die Erforschung von Literatur- und Kulturgeschichte*, Berlin 2015, 233–245.
- Piez, Wendell, „Hierarchies within range space. From LMNL to OHCO“, in: *Balisage. The Markup Conference. Proceedings* (2014), <http://www.balisage.net/Proceedings/vol13/html/Piez01/BalisageVol13-Piez01.html>.
- Rogers, Gretchen L., *Zur Textgeschichte der Gedichte Goethes*. Diss., Baltimore, MD, 1938.
- Scheibe, Siegfried, „Von der Entstehungsgeschichte, der Textgeschichte und der zeitgenössischen Wirkungsgeschichte“, in: Siegfried Scheibe (Leitung)/Waltraud Hagen/Christel Laufer u.a. (Hg.), *Vom Umgang mit Editionen. Eine Einführung in Verfahrensweisen und Methoden der Textologie*, Berlin 1988, 160–204.
- Scheibe, Siegfried, *Kleine Schriften zur Editionswissenschaft*, Berlin 1997.
- Schöne, Albrecht, *Johann Wolfgang Goethe. Faust. Kommentare*, Frankfurt a. M. 2017.
- Szondi, Peter, „Über philologische Erkenntnis“, in: Ders., *Hölderlin-Studien. Mit einem Traktat über philologische Erkenntnis*, Frankfurt a. M. 1967, 9–34.
- Woessler, Winfried, „Althilologische editorische Terminologie in der Neuphilologie“, in: Gunter Martens (Hg.), *Editorische Begrifflichkeit. Überlegungen zu einem „Wörterbuch der Editionsphilologie“*, Berlin/Boston 2013, 93–102.
- Zeller, Hans, „Befund und Deutung. Interpretation und Dokumentation als Ziel und Methode der Edition“, in: Gunter Martens/Hans Zeller (Hg.), *Texte und Varianten*, München 1971, 45–89.

- Zeller, Hans, „Fünfzig Jahre neugermanistischer Edition. Zur Geschichte und künftigen Aufgaben der Textologie“, in: *Editio* 3 (1989), 1–17.
- Zeller, Hans, „Struktur und Genese in der Editorik. Zur germanistischen und anglistischen Editionsforchung“, in: *LiLi* 19/20 (1975), 105–126.
- Zeller, Hans, „Textologie und Textanalyse. Zur Abgrenzung zweier Disziplinen und ihrem Verhältnis zueinander“, in: *Editio* 1 (1987), 145–158.
- Zeller, Hans, „Zur gegenwärtigen Aufgabe der Editionstechnik. Ein Versuch, komplizierte Handschriften darzustellen (1959)“, in: Rüdiger Nutt-Kofoth (Hg.), *Dokumente zur Geschichte der neugermanistischen Edition*, Tübingen 2005, 194–214.

Online-Ressourcen

- GitHub-Issue *Search solution for words with inline markup*, <https://github.com/faustedition/faust-web/issues/196>.
- Jean-Paul-Portal Würzburg – Projekt *Hesperus*, http://www.jean-paul-portal.uni-wuerzburg.de/neue_werkausgabe/editorenwerkstatt_hesperus/.
- Metricalizer*, <https://metricalizer.de/de/about/>.
- Nestorchronick (Povest' vremennyx let)*, <http://pvl.obdurodon.org/browser.xhtml>.
- TEI-Guidelines*, <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Digital Genetic Editions. Towards Macroanalysis across Versions

Dirk Van Hulle

The focus on big data, distant reading and macroanalysis in Digital Humanities seems to have the immediate effect that close reading is forced into an antonymous position and non-digital literary studies or *Literaturwissenschaft* suddenly look provincial in comparison. But not all traditional forms of literary studies are microscopic or focused on close reading, and vice versa, not all digital forms of literary studies are macroscopic or panoramic. Distant reading can also be reductive in some ways, as it usually limits its ‘reading’ to only one version of texts.

Instead of working with this dichotomy between micro and macro, it might be more useful to think in terms of a continuum. Situated somewhere on this continuum between micro and macro is computer-assisted genetic criticism, the study of the dynamics of creative processes. In some respects, genetic criticism is concerned with the most microscopic aspects of literature: it involves the transcription of every single comma, every metamark, every letter in a manuscript, whether crossed out or not. But it does not limit itself to this microgenetic scrutiny. It also involves macrogenetic analyses, investigating the development of passages across versions, which is enabled by developments in automatic or computer-assisted collation.

In this article I would like to focus on variants made by the author her/himself and make a plea for a reevaluation of something that is often excluded from the literary canon, namely the things that never made it into the final texts. When it comes to scholarly editing, we tend to prioritize the published works. A writer’s ‘complete works’ are usually arranged according to the author’s published works.

D. Van Hulle (✉)
University of Antwerp, Antwerp, Belgium
E-Mail: Dirk.Vanhulle@ell.ox.ac.uk

Scholarly editing has traditionally focused on establishing critically edited texts. The reading texts of an author's works understandably constitute the core of the critical edition, while notes and drafts are generally only mentioned insofar as they provide evidence for the establishment of the reading text. These manuscripts and other relevant materials (such as letters and diaries) have been dubbed the 'grey canon' by S. E. Gontarski.¹ But even this reevaluation suggests a hierarchy between the 'real' canon and the 'grey' canon. Items from this grey canon are regarded (and treated) as mere satellites orbiting around the central planet of the canon of the published works.

But sometimes writers keep a notebook containing ideas and loose jottings that did not necessarily lead to any particular work. Some of the notes will turn out to be dead ends, others will end up in various works. To fully understand the dynamics of the writing process, it is necessary to build a digital infrastructure that organizes an author's works not only (sect. 1) according to the canon, but also (sect. 2) in a different way, which will be referred to as a 'dysteleological approach'. The model I would like to propose includes both these approaches.

1. The teleological approach organizes the genetic edition according to the logic of the author's canon and its *avant-texte*, treating the work as an *œuvre*, consisting of a set of separately published works. Some of the digital tools that support this approach are (sect. 1.1) a collation engine; (sect. 1.2) a system to trace cuts and notes in the writing process; and (sect. 1.3) a set of statistics applied to information in the XML encoding of the transcriptions. The question that will be investigated is what these may yield for literary studies.
2. The dysteleological approach organizes the genetic edition according to the logic of the continuous process of writing, regarding the work as *travail* – the hard work that goes into writing and that does not always necessarily make it into publication. This view corresponds to Paul Valéry's image of the snake, which keeps moving 'on', shedding its skin once in a while. Whereas the teleological approach focuses on the metaphorically shed skins (the published works), the dysteleological approach focuses on the snake's vestigial organs and follows the movements of the snake as it proceeds without necessarily having a clear goal. Some of the digital tools that support this approach are (sect. 2.1) the organization of the genetic edition according to the logic of the notebook; (sect. 2.2) the search engine, which searches across works and across versions, not only for text, but also for visual elements such as doodles; and (sect. 2.3) the author's digitalized library.

¹S. E. Gontarski, "Greying the Canon: Beckett in Performance," in: S. E. Gontarski/Anthony Uhlmann (eds.), *Beckett after Beckett*, Gainesville 2005, 141–157, here: 143).

1 A Teleological Approach

A ‘complete works’ edition is typically organized according to an author’s canon: each of the author’s published novels serves as an endpoint; the manuscripts, letters, diary entries are mentioned to the extent that they lead up to this ‘telos’. This is a perfectly legitimate approach, but it does raise the question: What belongs to an author’s canon?

For instance, in the case of the bilingual Irish author Samuel Beckett: What is the Beckett canon? In 2001, Ruby Cohn published *A Beckett Canon*,² in which she discusses mainly the published, but also a few unpublished ones. In the meantime, we have found some new unpublished works, which raises the issue whether they belong to the canon or not. Should they be included in a ‘Complete Works’ edition? How complete is a complete works edition without them?

In Beckett’s case, they do deserve a place in the digital edition, called the *Beckett Digital Manuscript Project (BDMP)*.³ The digital infrastructure of the BDMP is organized in such a way that, wherever the user happens to be in the edition, s/he can always choose any sentence and compare it to all the other extant versions of that sentence in a synoptic sentence view. This option arranges the multiple versions in chronological order and enables the user to compare the versions (‘versioning’). For instance, in Beckett’s play, *Krapp’s Last Tape*, there is a moment which, in a cinematic context, would be called a ‘continuity error’. In all the Faber and Faber editions, the protagonist Krapp listens to an old tape about his mother, ‘a-dying, *in the late autumn*, after her long viduality’ (emphasis added); he winds back the tape, and then hears ‘a-dying, after her long viduality’ (without ‘in the late autumn’), which is clearly an error (*BDMP3*, ET5, 4r; ETC, 4r). The synoptic sentence view enables readers to compare all the extant versions in both the English texts and in Beckett’s own French translation.

1.1 Collation Engine

In addition to this form of bilingual ‘versioning’, the *BDMP* enables users to collate sentences in either French or English manuscripts and editions, by activating the automatic collation tool powered by *CollateX* (developed by Ronald Haentjens Dekker), which compares the sentences and highlights all the variants between them. In the case of the continuity error mentioned above, the digital collation clearly shows the moment in the genesis where the error occurred.⁴

²Ruby Cohn, *A Beckett Canon*, Ann Arbor 2001.

³Samuel Beckett, *Krapp’s Last Tape/La Dernière Bande. A Digital Genetic Edition* (Series ‘The Beckett Digital Manuscript Project’, module 3), eds. Dirk Van Hulle/Vincent Neyt, Brussels 2015, <http://www.beckettarchive.org> (retrieved on May 20, 2018), abbreviated as *BDMP3*.

⁴Cf. *BDMP3*, <http://www.beckettarchive.org/krapp/comparesentences/0186> (retrieved on May 20, 2018).

Spotting differences between versions of a text seems like a job a computer should be able to do quite easily. In practice however, in most cases an apparatus created automatically by the collation software currently available (such as *CollateX*, *HyperCollate*, *Juxta*, Multi Version Documents and the TEI-Comparator) does not yet match up to a critical apparatus created by hand by an editor. This is especially the case with modern manuscripts as they contain in-text variation (such as additions and open variants). There are conflicting opinions on how best to encode these texts with a view to collation,⁵ as well as on the scholarly validity of automatic collation output. Computer-assisted collation has a relatively long tradition in digital humanities, going back at least to the use of *TUSTEP* by Hans Walter Gabler and his team for the production of their edition of James Joyce's *Ulysses*. So far, collation tools have always been used as tools for the *editor*, in order to produce a critical apparatus. The model I propose offers automatic collation as a tool for the *user* to highlight variants. We might call it a 'collation engine', by analogy with a 'search engine'.

The collation engine takes the digital transcriptions as input and performs a service for the user, who can leave certain witnesses out of the collation if they so choose. In this way, instead of turning the critical apparatus into the most tedious part of a critical edition, a digital archive can offer automatic collation as an alternative tool to help users discover complex and therefore interesting textual instances in the manuscripts and other textual versions. The integration of *CollateX* as a collation engine in the *BDMP* has the advantage for genetic critics that it can collate manuscript versions: deletions and additions can be recognized and presented as such in the collation output.

The rationale behind offering this tool to the reader rather than only to the editor is that the collation of different versions is an active form of reading across versions, which is neither a privilege nor a chore that is exclusively reserved for textual scholars. At this moment collation tools may not yet produce results that are 100 % reliable (i.e. not as reliable as a critical apparatus), but the same goes for other everyday tools such as search engines, which also produce results that have to be filtered by the reader. While we continue to collaborate to make the collation algorithm smarter, reading across versions with a collation engine does not require more ingenuity from readers than a search engine does. The advantage is that textual variants do not need to be relegated to the so-called 'Variantenfriedhof' of a critical apparatus, but that readers can actively engage with the differences between versions, in a way that allows them to zoom in and out and thus move freely on the continuum between close and slightly more distant reading.

⁵Barbara Bordalejo, "The Encoding System," in: Prue Shaw (ed.), *The Commedia of Dante Alighieri. A Digital Edition*, Saskatoon 2010; Peter Robinson, "Towards a Theory of Digital Editions," in: *Variants* 10 (2012), 105–131; Desmond Schmidt/Robert Colomb, "A data structure for representing multi-version texts online," in: *International Journal of Human-Computer Studies* 67/6 (2009) 497–514; Desmond Schmidt, "The role of markup in the digital humanities," in: *Historical Social Research* 37 (2010), 125–46.

1.2 Cuts and Notes

The digital infrastructure that enables sentence comparison is based on a simple numbering of sentences, linked to the first edition as anchor text. If a user chooses a sentence that was cut during the process of revision, this sentence receives the sentence number of the previous sentence (e.g. <seg n="251">) that *did* make it into the published version, followed by an extra number (e.g. <seg n="251|001"). In the synoptic sentence view, this cut sentence (or series of cut sentences) is visualized in bold, following its anchor sentence (e.g. anchor sentence 251, followed by sentences 251|001, 251|002, ... in bold). This way, the cuts and the author's process of creative undoing can be mapped.

But even so, the edition may, at this moment, be still too teleologically organized. On the opening page, the menu invites the reader/user to choose a work before they can enter the genetic dossier. This is the *avant-texte* approach to genetic criticism: by speaking of an *avant-texte*, the default assumption is that there is a text in the first place. But, of course, there are also notes and drafts that did *not* lead to a published text.

Therefore, it is useful to add an option for users to start from the notebooks. Instead of only incorporating those parts of a notebook that show the drafts of a particular work that made it into the published version, this 'notebooks' option (in the edition's general menu of options) shows the notebook in its entirety, including those parts that never made it into publication.

1.3 Distant Genetic Reading

If we take Moretti's original definition of distant reading as a form of *indirect* reading, not necessarily making use of computers, the distance from the text is linked to the "ambition" of the research, according to Moretti.⁶ Ambition comes from the Latin verb 'ambire', 'going around' to solicit votes, hence the notion of flattery and a thirst for honour, favour and popularity. It is symptomatic of our digital age that this originally pejorative term has become such a central concept in academia. In literary studies, the question is whether textual scholarship can connect to the more ambitious going 'around' of distant reading and still go 'on' doing what it has always been good at.

⁶In his essay "Conjectures on world literature" (2000), Moretti related distant reading to making use of secondary literature rather than primary literature: "literary history will become very different from what it is now: it will become 'second hand': a patchwork of other people's research, without a single direct textual reading. Still ambitious, and actually even more so than before [...] but the ambition is now directly proportional to the distance from the text: the more ambitious the project, the greater must the distance be." (Franco Moretti, *Distant Reading*, London/New York 2013, 48).

Tab. 1 Deleted, added and modified words in the manuscripts

Molloy	5 % deleted, 2 % added, 1 % modified
Malone meurt/Malone Dies	7 % deleted, 3 % added, 1 % modified
En attendant Godot/Waiting for Godot	7 % deleted, 2 % added, 0.1 % modified
L'Innommable/The Unnamable	6 % deleted, 3 % added, 1 % modified
Fin de partie/Endgame	5 % deleted, 4 % added, 1 % modified
Krapp's Last Tape/La Dernière Bande	5 % deleted, 5 % added, 1 % modified
Stirrings Still/Soubresauts	14 % deleted, 5 % added, 1 % modified
Comment dire/what is the word	15 % deleted, 9 % added, 0.01 % modified

A modest form of indirect or distant genetic reading is already being practiced by making use of the XML encoding of manuscript transcriptions. The and <add> tags can be usefully deployed to calculate the percentage of words that remained 'stable' across versions, how many were cut and how many added. This form of indirect reading allows us, for instance, to investigate an author's poetics. In the example of Samuel Beckett, the percentages relating to the geneses of eight works have been calculated and visualized in pie charts in the Beckett Digital Manuscript Project.⁷

This is a form of distant reading applied not to one published version of many texts – as is usually the case in distant reading projects so far – but to the complete genetic dossier of a set of texts. Even though the set is admittedly 'unambitious', this form of indirect reading has already yielded some interesting results for literary criticism, notably for the study of the author's poetics. Whenever Beckett was asked about his poetics, he almost automatically took James Joyce as his point of reference, which is understandable since early in Beckett's career Joyce had been a sort of mentor, and certainly a role model, to the young writer. But this also meant that Beckett needed to take a distance from Joyce in order to find his own voice. While Joyce was always adding to his manuscripts, Beckett's method consisted of undoing, as he claimed in an interview with James Knowlson: "I realised that Joyce had gone as far as one could in the direction of knowing more, [being] in control of one's material. He was always adding to it; you only have to look at his proofs to see that. I realised that my own way was in impoverishment, in lack of knowledge and in taking away, subtracting rather than adding."⁸

The pie charts allow us to compare this statement to the statistics of deleted, added and modified words in the manuscripts (Tab. 1).

The overall pattern is that Beckett, indeed, cut more than he added, and towards the end of his career, this general trend becomes more outspoken. The ratio remains relatively stable (on average 1 added word for every 3 deleted words), but

⁷ <http://www.beckettarchive.org/stats.jsp> (retrieved on May 20, 2018).

⁸ James Knowlson, *Damned to Fame. The Life of Samuel Beckett*, London 1996, 352.

the results so far show more textual instability towards the end of Beckett's career than in the years shortly after the Second World War. Paradoxically, the more experience the author gained as a writer the more hesitant his writing became. The result of this indirect reading is modest in scope, but it does indicate how textual scholarship can contribute to new forms of reading in innovative ways by enabling forms of indirect reading applied to more than one version of a text, including its genesis.

2 A Dysteleological Approach

The system to chart cuts in the manuscripts (cf. Sect. 1.2 above) is designed from a teleological perspective: the sentences that did not make it into the published text are numbered with reference to the closest preceding sentence that did make it to the 'telos'. But we also need to retrace the creative process from a non-teleological or 'dysteleological' perspective. The above description of the statistics may give us too neat an impression of an evolution in Beckett's writing. Because the numbering of the sentences can only be applied to manuscripts that already show more or less a correspondence with the published work, the computation of data excludes the fragments that are hard to classify. For instance, in the period between the publication of *En attendant Godot* and the writing of *Fin de partie*, Beckett wrote several short dramatic fragments that contain both stylistic elements recalling *Godot* and aspects that prefigure scenes in *Fin de partie*. On one of these fragments, Beckett retrospectively wrote 'Avant *Fin de partie*'. We could therefore label all these fragments 'Before Endgame', but then we need to make a distinction between 'before' in the sense of 'leading up to the first drafts of *Endgame*' and 'before' in the sense of 'not yet belonging to the *avant-texte* of *Endgame*'.

The latter category contains abandoned fragments (AF). The sentences in these fragments are numbered and marked with the abbreviation AF (e.g. <seg n="AF011001">). The statistics function (see sect. 1.3 above) can be of help in determining the position of the abandoned fragment: it strips all the tags from the encoding, replacing the add and del tags by a special code; the first and the last word of an abandoned fragment are marked; the data are stored in a separate file, keeping a record of the total number of words in the version, the position of the first word and that of the last word of the AF. In this particular case, the total number of words in the longest version of *Fin de partie/Endgame* is 19,600; the total number of words in the early version UoR1660 is 16,439 (83.88 %); the first abandoned fragment in this version (AF1) ranges from AF10695 to AF11061. With these data, the position of the abandoned fragment can be located and visualized, for instance on a strip representing the entire text of UoR1660, 1.87 % of which represents the AF, located in the second half of the draft (after 54.57 % of the text).

In a teleological approach, none of these abandoned fragments is taken into account in the statistics relating to *Endgame*, because it is unclear if they can be

included in the genesis of this play. If we do include them – arguing that Beckett apparently needed this period of trial and error to find the right themes and shape for his play – this complicates the all too neat computationally generated pattern (from relative decisiveness to more hesitation) outlined above under 1.3. This is not to suggest that computational methods are incapable of dealing with these more complicated aspects of literary geneses, but it does imply that we need to think of ways to also accommodate a dysteleological approach to the writing process. This dysteleological approach requires that the infrastructure and architecture of the digital edition is not dictated solely by the logic of the published text but also takes the logic of the notebook into account.

2.1 The Logic of the Notebook

An interesting case to study this logic is the genesis of James Joyce’s *Finnegans Wake*.⁹ About fifty years ago, A. Walton Litz published his essay “Uses of the *Finnegans Wake* Manuscripts”, noting that “the published texts of *Finnegans Wake* are *corrupt* in many places”. His conclusion was:

Of course, any editor must be cautious when he goes beyond Joyce’s own corrections and emends from the manuscripts; Joyce’s ‘love of accidentals’ is well known, and some of the apparent errors may have received his silent sanction. But there is no reason why a good critical text of *Finnegans Wake* should not ultimately be produced, based upon judicious use of the manuscripts.¹⁰

This “judicious use” was facilitated in the 1970s when Garland published the facsimile edition of Joyce’s manuscripts (the *James Joyce Archive* in more than 60 volumes).¹¹ But even this facsimile edition of the manuscripts is arranged according to the book’s final structure.

As a consequence, a pivotal document in the creative process, the so-called ‘red-backed’ or ‘Guiltless’ copybook,¹² has been published in separate volumes. In this facsimile edition, it is not possible to take one volume of the *James Joyce Archive* to study the document that is preserved at the British Library under catalogue number MS Add. 47471b. Since the copybook contains drafts of various chapters, the facsimiles have been arranged in four different volumes.¹³ And

⁹James Joyce, *Finnegans Wake*, London 1939.

¹⁰Walton A. Litz, “Uses of the *Finnegans Wake* Manuscripts,” in: Jack P. Dalton/Clive Hart (eds.), *Twelve and a Tilly. Essays on the Occasion of the 25th Anniversary of *Finnegans Wake**, Evanston 1965, 99–106, here: 100.

¹¹James Joyce, *James Joyce Archive*, eds. Michael Groden, Hans W. Gabler, David Hayman et al., New York 1977–1979.

¹²*British Library*, BL MS 47471b, named after the first word on its first page.

¹³*James Joyce Archive*, vols. 45, 46, 47, and 48.

because sometimes a page contains parts of different sections, this page had to be reproduced twice (or even three times in the case of page 30r).

In the meantime, what Litz called “judicious use of the manuscripts” is taking different shapes in the digital age. Around the same time as the publication of the *James Joyce Archive*, Ted Nelson coined the concept of ‘transclusion’ in his book *Literary Machines*,¹⁴ and before coining the term he actually already formulated the idea behind transclusion in his 1965 description of hypertext, describing it as “the same content knowably in more than one place.”¹⁵ After more than fifty years, this idea still serves as a useful concept to make the teleological and dysteleological approaches interact to facilitate the reconstruction of the creative process. The logic of the notebook is that it has a materiality (a physical shape and size) of its own, which usually does not coincide with the length of the work that is being written. Still, these physical confines of the notebook sometimes do have an influence on the writing process. In the case of the ‘Guiltless’ copybook, for instance, Joyce tried to stay within the confines of this creative space for as long as possible, eventually ending up writing on every blank space he could find, filling them even in retrograde direction until the notebook was completely filled with text. The well-known effect of the ‘text produced so far’ as a source of inspiration to move on was such that Joyce made use of it to the full by trying to avoid having to start a new notebook and be confronted with the first blank page of an object that he had not yet made his own.

Similarly, Samuel Beckett wrote his play *En attendant Godot* in a notebook that he did not wish to part with until the end of his life. The text is written on the right-hand pages until he reached the end of the notebook, after which Beckett started filling the blank verso pages, rather than continuing the text in a new notebook. This suggests that the notebook as a physically confined unit plays a role in the dynamics of the writing process.

At the same time, it is possible and useful to confront this logic of the notebook with the logic of the ‘telos’ to see how one document may contain more than one version of more than one section. The succession of pages is a chronology that differs from that of the succession of versions. A transclusive rearrangement allows us to give shape to these two temporal axes.

An editorial approach inspired by digital poetics or digital genetics can try to accommodate both a teleological and a dysteleological approach by treating the notebook as a pivot between exogenesis (all aspects of the genesis relating to

¹⁴Theodor H. Nelson, *Literary Machines*, Sausalito 1981.

¹⁵Theodor H. Nelson, “A File Structure for the Complex, the Changing and the Indeterminate,” in: *Proceedings of the ACM 20th National Conference*, New York 1965, 84–100, <https://doi.org/10.1145/800197.806036>; cf. also Theodor H. Nelson/Robert A. Smith, “Back to the Future. Hypertext the Way It Used to Be,” in: *Proceedings of the ACM 18th Conference on Hypertext and Hypermedia*, New York 2007, 227–228, <http://xanadu.com/XanaduSpace/btf.htm> (retrieved on May 20, 2018).

external source texts), endogenesis (the ‘internal’ writing process) and epigenesis (the continuation of the genesis after publication), that is, as the place where the author creates a mental space to develop ideas and extract elements from other sources, creatively undoing the original context, and appropriating or processing the notes in a new composition.

2.2 The Search Engine

The advantage of being able to search through several *avant-textes* adds an extra dimension to traditional editions. It allows readers to follow the chronological development of a particular concept throughout an author’s *œuvre*, but also its development throughout the writing process of each work, thus creating a double temporal axis to search across works and across versions. Not only textual units, also visual aspects of the manuscripts can be searched, such as stage drawings, diagrams and doodles. The doodles can be categorized according to a typology, divided into four main categories: objects, organisms, shapes and symbols. The encoding also facilitates other searches such as the systematic search for dates, gaps (blank spaces in manuscripts), calculations and intertextual references. The latter category partially links up with the author’s personal library.

2.3 Integration of the Writer’s Digitalized Library

As noted above (2.1), notebooks with reading notes function as a pivot between exogenesis, endogenesis and epigenesis. For that reason, it may be useful to include the writer’s library in a genetic edition if it is still extant, and – to the extent that it is no longer extant – to reconstruct a virtual library. For the reconstruction of the virtual library, text re-use software such as Tracer can be deployed. In the analysis of intertextual relations, a writer’s notebooks and drafts can play a pivotal role. The emphasis in this kind of analysis is on the transformative operation, the processing of intertexts. Daniel Ferrer suggests an approach that focuses on the intertextual *in operation* (“le fonctionnement intertextuel”) rather than on the intertext itself.¹⁶

Ever since the term “intertextuality” was coined by Julia Kristeva in 1966 it has been made to mean numerous different things by various researchers, to such an extent that Kristeva herself took a distance from the notion less than a decade after she had introduced it.¹⁷ Building on Bakhtin’s notion of ‘dialogism’, Kristeva argued that any text is “constructed as a mosaic of quotations; any text is the

¹⁶Daniel Ferrer, “Quelques remarques sur le couple intertextualité-genèse,” in: Paul Gifford/Marion Schmid (eds.), *La création en acte*, Amsterdam 2007, 205–216, here: 206.

¹⁷Julia Kristeva, *La Révolution du langage poétique*, Paris 1974, 59–60.

absorption and transformation of another”.¹⁸ Similarly, Roland Barthes stated that every text “belongs to the intertextual”.¹⁹ For Kristeva ‘intertextuality’ denoted the way (not just literary) texts emerge from a particular semiotic order, involving complex semiotic presuppositions.²⁰ In “Presupposition and Intertextuality”, Jonathan Culler noted that intertextuality is “less a name for a work’s relation to particular prior texts than a designation of its participation in the discursive space of a culture”.²¹ In the 1980s and 1990s, theorists such as Gérard Genette and Michael Riffaterre recalibrated the notion of intertextuality, Genette by framing it as a relation of “co-presence”²² in an elaborate categorization of what he termed “transtextuality”, and Riffaterre by insisting that intertextuality is less a matter of writing than a matter of reading.

Even though, from Julia Kristeva’s perspective, the term ‘intertextuality’ was misappropriated and applied to what she called a “banal” understanding of the term in the sense of “study of sources” [“critique des sources”],²³ there are understandable reasons for this phenomenon of misappropriation in literary criticism. One of these is that many of the traditional terms to denote textual interrelations (such as ‘source’, ‘influence’, or ‘borrowing’) imply a hierarchy, suggesting that the later text is derivative and unoriginal.²⁴ ‘Intertextuality’ has the advantage that it does not suggest this inequality.

This is not an exclusively literary matter. Michael Baxandall’s criticism of the term ‘influence’ (and implicitly of the term ‘source’) is still relevant, to both art history and literary studies. “Influence”, he writes, is “a curse of art criticism primarily because of its wrong-headed grammatical prejudice about who is the agent and who the patient”.²⁵ With reference to Joyce Studies, Scarlett Baron succinctly summarized the situation at the beginning of the twenty-first century as “a choice between two contending paradigms: influence on the one hand, intertextuality on the other”.²⁶

A recent exhibition of Rubens’ ‘intertextual’ appropriations in Frankfurt (the *Städel Museum*) also showed how the artist gradually processed his own impression and sketch of a sculpture of a Centaur and then used it to give shape to a painting

¹⁸ Julia Kristeva, *Desire in language. A semiotic approach to literature and art*, ed. Leon Roudiez, New York 1980, 66.

¹⁹ Roland Barthes, “From Work to Text,” in: J. V. Harari (ed.), *Textual Strategies*, Ithaca 1979, 73–81, here: 77.

²⁰ Kristeva (Ann. 18), 37 f.

²¹ Jonathan Culler, *The Pursuit of Signs. Semiotics, Literature, Deconstruction*, Ithaca 1981, 103.

²² Genette, Gérard, *Palimpsestes. La littérature au second degré*, Paris 1982, 8.

²³ Kristeva (Ann. 17), 59.

²⁴ Gregory Machacek, “Allusion,” in: *PMLA* 122/2 (2007), 522–536, here: 525.

²⁵ Michael Baxandall, *Patterns of Intention. On the Historical Explanation of Pictures*, New Haven 1985, 58–59.

²⁶ Scarlett Baron, ‘Strandentwining Cables’. *Joyce, Flaubert, and Intertextuality*, Oxford 2012, 8.

of a religious theme, Pilate presenting Christ to the crowd ('Ecce Homo'). From this 'intertextual' perspective the painting acquires several layers of complexity. Not only does it show Christ with a Centaur's torso, but the whole transformation of a Greek mythological theme into a Christian topic is a complex process, in which the artist's drafts or sketches play a crucial intermediary role. Similarly, in literary studies, a notebook is a pivot between exogenesis and endogenesis. In many cases this means its notes derive from multiple sources and inform not just one but multiple works by the author. In that sense, the notebook invites researchers to read an author's works from a certain distance and see them as a continuum. This panoramic genetic reading enables scholars to study not just a work in progress but an *œuvre* in progress, and as soon as more digital genetic editions are available, perhaps even entire literary periods in progress, including macroanalyses across versions.

3 Conclusion: Towards Macroanalyses Across Versions

In the last few years, new developments in digital humanities, notably in the field of 'text re-use' software, have made the influence/intertextuality debate more relevant than ever. In February 2018, Dennis McCarthy and June Schlueter reported on the results of their research that use plagiarism software on Shakespeare's works, noting several correspondences between a 1576 manuscript by George North and more than twenty passages in Shakespeare's texts.²⁷ Precisely against the background of these digital developments it is important to reassess the notion of 'intertextuality', because the nuances in the 'influence'/'intertextuality' debate are not always taken into account in Digital Humanities, as for instance in the title 'Measuring the Influence of a Work by Text Re-Use'.²⁸ Matthew Jockers' fascinating macroanalysis of stylistic and thematic information from the books in the Literary Lab corpus is headed by the chapter title "Influence" whereas the term 'intertextuality' occurs in a sentence that is related to microanalysis: "Attempts to demonstrate literary imitation, intertextuality, and influence have relied almost entirely upon close reading."²⁹ It seems important that digital text analysis pays renewed attention to the 'influence' versus 'intertextuality' debate, and here I see an important theoretical avenue for *Digitale Literaturwissenschaft*.

²⁷ Alison Flood, "Plagiarism software pins down new source for Shakespeare's plays," in: *The Guardian* (9 February 2018).

²⁸ Marco Büchler/Anette Geßner/Monica Berti et al., "Measuring the Influence of a Work by Text Re-Use," in: Stuart Dunn/Simon Mahony (eds.), *Digital Classicist Supplement. Bulletin of the Institute of Classical Studies*, Wiley-Blackwell 2013.

²⁹ Matthew L. Jockers, *Macroanalysis: Digital Methods and Literary History*, Urbana/Chicago/Springfield 2013, 156.

Riffaterre defined intertextuality as the *reader's* perception of the links between one work and others, which preceded or followed it (“la perception par le lecteur, de rapports entre une œuvre et d’autres qui l’ont précédée ou suivie”³⁰). For the purposes of studying intertextuality in a digital context, we can build on this definition, but not without readjusting an important detail: ‘the’ reader is a generalization which, by means of including the digital reconstruction of an author’s library in a digital scholarly edition, can be replaced by a very concrete reader: the author as reader, a reader who was also a writer. The advantage of this type of reader is that they often leave numerous traces of their reading (in the form of marginalia and notes) and that we consequently have quite a lot of research data.

Still, ‘a lot’ is not enough. Matthew Jockers noted that “we have reached a tipping point, an event horizon where enough text and literature have been encoded to both allow and, indeed, force us to ask an entirely new set of questions about literature and the literary record”.³¹ My sense is that, in digital genetic editing, there is still a long way to go before panoramic reading of entire periods in progress and macroanalyses across versions will be effective. That is why it is so important that, as Matthew Jockers notes, macroanalysis “is not offered as a challenge to, or replacement for, our traditional modes of inquiry” but as a complement.³² The digital genetic *microanalysis*, involving careful transcription of barely legible manuscripts and the encoding of deletions, additions and marginalia, is the key to making future genetic *macroanalyses* possible.

References³³

- Baron, Scarlett, *‘Strandentwining Cables’. Joyce, Flaubert, and Intertextuality*, Oxford 2012.
- Barthes, Roland, “From Work to Text,” in: J. V. Harari (ed.), *Textual Strategies*, Ithaca 1979, 73–81.
- Baxandall, Michael, *Patterns of Intention. On the Historical Explanation of Pictures*, New Haven 1985.
- Beckett, Samuel, *Krapp’s Last Tape/La Dernière Bande. A Digital Genetic Edition* (Series ‘The Beckett Digital Manuscript Project’, module 3), eds. Dirk Van Hulle/Vincent Neyt, Brussels 2015, <http://www.beckettarchive.org>, abbreviated as *BDMP3*.
- Bordalejo, Barbara, “The Encoding System,” in: Prue Shaw (ed.), *The Commedia of Dante Alighieri. A Digital Edition*, Saskatoon 2010.
- Büchler, Marco/Geßner, Anette/Berti, Monica et al., “Measuring the Influence of a Work by Text Re-Use,” in: Stuart Dunn/Simon Mahony (eds.), *Digital Classicist Supplement. Bulletin of the Institute of Classical Studies*, Wiley-Blackwell 2013.
- Cohn, Ruby, *A Beckett Canon*, Ann Arbor 2001.
- Culler, Jonathan, *The Pursuit of Signs. Semiotics, Literature, Deconstruction*, Ithaca 1981.

³⁰Michael Riffaterre, “La trace de l’intertexte,” in: *La Pensée* 215 (1980), 4.

³¹Jockers (Ann. 29), 4.

³²*Ibid.*, 171.

³³All digital references were last viewed on May 20, 2018.

- Ferrer, Daniel, "Quelques remarques sur le couple intertextualité-gènèse," in: Paul Gifford/Marion Schmid (eds.), *La création en acte*, Amsterdam 2007, 205–216.
- Flood, Alison, "Plagiarism software pins down new source for Shakespeare's plays," in: *The Guardian* (9 February 2018).
- Genette, Gérard, *Palimpsestes. La littérature au second degré*, Paris 1982.
- Gontarski, S. E., "Greying the Canon: Beckett in Performance," in: S. E. Gontarski/Anthony Uhlmann (eds.), *Beckett after Beckett*, Gainesville 2005, 141–157.
- Jockers, Matthew L., *Macroanalysis: Digital Methods and Literary History*, Urbana/Chicago/Springfield 2013.
- Joyce, James, *Finnegans Wake*, London 1939.
- Joyce, James, *James Joyce Archive*, eds. Michael Groden/Hans W. Gabler/David Hayman et al., New York 1977–1979.
- Knowlson, James, *Damned to Fame. The Life of Samuel Beckett*, London 1996.
- Kristeva, Julia, *Desire in language. A semiotic approach to literature and art*, ed. Leon Roudiez, New York 1980.
- Kristeva, Julia, *La Révolution du langage poétique*, Paris 1974.
- Litz, A. Walton, "Uses of the Finnegans Wake Manuscripts," in: Jack P. Dalton/Clive Hart (eds.), *Twelve and a Tilly. Essays on the Occasion of the 25th Anniversary of Finnegans Wake*, Evanston 1965, 99–106.
- Machacek, Gregory, "Allusion," in: *PMLA* 122/2 (2007), 522–536.
- Moretti, Franco, *Distant Reading*, London/New York 2013.
- Nelson, Theodor H., "A File Structure for the Complex, the Changing and the Indeterminate," in: *Proceedings of the ACM 20th National Conference*, New York 1965, 84–100, <https://doi.org/10.1145/800197.806036>.
- Nelson, Theodor H., *Literary Machines*, Sausalito 1981.
- Nelson, Theodor H./Smith, Robert A., "Back to the Future. Hypertext the Way It Used to Be," in: *Proceedings of the ACM 18th Conference on Hypertext and Hypermedia*, New York 2007, 227–228, <http://xanadu.com/XanaduSpace/btf.htm>.
- Riffaterre, Michael, "La trace de l'intertexte," in: *La Pensée* 215 (1980), 4.
- Robinson, Peter, "Towards a Theory of Digital Editions," in: *Variants* 10 (2012), 105–131.
- Schmidt, Desmond, "The role of markup in the digital humanities," in: *Historical Social Research* 37 (2010), 125–46.
- Schmidt, Desmond/Colomb, Robert, "A data structure for representing multi-version texts online," in: *International Journal of Human-Computer Studies* 67/6 (2009) 497–514.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Annotation als Markup *avant la lettre*

Jan Christoph Meister

1 **Status quo: Das methodologische Delta**

In der Textthermeneutik als zentralem und für das traditionelle Selbstverständnis der Philologien wesentlichem Arbeitsbereich spielte der Einsatz digitaler Verfahren bislang kaum eine Rolle.¹ Zwar profitiert auch die literaturwissenschaftliche Auslegung von Textbedeutungen mittelbar durchaus vom *Digital Turn*: Software und webbasierte Kollaborationsumgebungen können den philologischen Arbeitsprozess von der Edition bis zur Lektüre des Einzeltextes unterstützen; Kollokationen und Datenbanken werden zu Recherchezwecken genutzt; Verfahren des *Data Mining*, der *Computational Stylometry* und des *Topic Modeling* bieten

¹Die nachfolgend skizzierte Bestandsaufnahme basiert u. a. auf bisherigen Erhebungen im Zeitraum 2014–2015, die im Rahmen von DARIAH-DE AP 5.4/Teilprojekt *Stakeholdergesellschaften* in Abstimmung mit dem Fachverband *DHd – Digital Humanities im deutschsprachigen Raum* (www.dig-hum.de [letzter Aufruf 8.1.2016]) vorgenommen wurden. Ziel dieses (2016 abgeschlossenen) Teilprojekts war „die Ermittlung der Bedarfe an digitalen Tools und Services, die sich aus der Perspektive von Fachwissenschaftlern konkret im Kontext ihrer je spezifischen Forschungsfragen wie Fachkulturen ergeben. Dieser nutzerzentrierte Ansatz soll sowohl zur Formulierung von Entwicklungsdesideraten und Use-Case-Szenarien beitragen wie das Interesse und die Akzeptanz der Wissenschaftler in den verschiedenen geisteswissenschaftlichen Fachcommunities für das Methodenparadigma des Big-Data-Approach fördern“ (Meister u. a., *DARIAH-DE Folgeantrag, Arbeitspaket 5.4*, internes Dokument 2013).

In den Beitrag sind Teile des DFG-LIS-Antrages *forTEXT* eingegangen. Die im Anhang beigefügte Sichtung und Evaluation von Annotationsplattformen wurde im Rahmen der Antragstellung von Marco Petris erarbeitet.

J. C. Meister (✉)
Universität Hamburg, Hamburg, Deutschland
E-Mail: jan-c-meister@uni-hamburg.de

neue Möglichkeiten für die Exploration großer Textkorpora mit Blick auf die Verbreitung und Genese inhaltlicher wie formaler Textmerkmale und eröffnen so nicht nur für die empirische Validierung von traditionell erarbeiteten Forschungshypothesen, sondern auch als Heuristiken neue Perspektiven, die für die Texthermeneutik durchaus relevant sein können. Aber wo es um die Textinterpretation als Bedeutungsauslegung im klassischen Sinne geht, finden computergestützte Ansätze bislang kaum Verwendung.²

Instrumentell bei vorbereitenden Verfahrensschritten durchaus in Anspruch genommen, ist das digitale Methodenparadigma dem philologischen Kerngeschäft der Texthermeneutik konzeptionell bislang suspekt geblieben. Diese Skepsis ist durchaus verständlich, stellt man die Spezifik des geisteswissenschaftlichen Wissensbegriffs selbst in Rechnung, der seit der Romantik durch zwei Leitideen bestimmt wird:

1. Historizität des Forschungsgegenstandes. – Das Erkenntnisobjekt ‚Text‘ wird in der hermeneutischen Praxis grundsätzlich als ein in stetig wandelbare soziale und kulturhistorische Kontexte eingebettetes Phänomen aufgefasst.
2. Diskursivität und Unabgeschlossenheit des hermeneutischen Erkenntnisprozesses. – Die Exploration und Auslegung von Textbedeutungen werden als rekursiv-iterative Prozesse praktiziert, die aufgrund der Dynamik ihrer Kontexte prinzipiell infinit sind; es gibt keinen hermeneutischen *Finite State* und keine absolut ‚wahre‘ Interpretation eines gegebenen Textes, sondern eine Bandbreite von mehr oder weniger plausiblen und anschlussfähigen Interpretationsvarianten.

Beide Leitideen, die das Selbstverständnis der Literaturwissenschaften nachhaltig geprägt haben, sehen traditionell arbeitende Philologen infrage gestellt von einer Technologie, deren Anwendung auf genuin hermeneutische Fragestellungen scheinbar zwingend einen reduktionistischen Begriff von hermeneutischer Erkenntnis als formal operationalisier- und technologisch (re-)produzierbarem Resultat voraussetzt.

²Von der grundsätzlichen Problematik zeugen nicht zuletzt die mehrfach gescheiterten Versuche, in den Geisteswissenschaften multifunktionale virtuelle Forschungsumgebungen zu etablieren, die mehr als nur eine vorab genau spezifizierte Aufgabenstellung (Texteditorik; Konkordanzerstellung etc.) bewältigen können. Weder die *bottom-up* ansetzende Strategie von Project Bamboo (<http://www.projectbamboo.org/> [letzter Aufruf 8.1.2016]), die von einer breiten Bedarfsanalyse und Befragung der geisteswissenschaftlichen Nutzercommunities ausging (vgl. zu den Problemen des Projekts Quinn Dombrowski, „What Ever Happened to Project Bamboo?“, in: *Literary and Linguistic Computing* 29/3 [2014], 326–339), noch die spezifischer bei den Bedarfen einer definierten Nutzercommunity ansetzenden Projekte wie etwa *NINES* (*Nineteenth-Century Scholarship Online*, <http://www.nines.org> [letzter Aufruf 8.1.2016]) oder die ambitionierten *top-down*-Strategien nationaler und multinationaler Infrastrukturvorhaben wie *DARIAH* haben bislang breite Akzeptanz gefunden.

Diese grundsätzliche Problematik wird dabei nicht selten noch verschärft durch die Fehleinschätzung von Potenzial und Anspruch gegenwärtiger digitaler Verfahren und Modellierungen. Die Schimäre ‚künstliche Intelligenz‘ spielt dabei ebenso eine Rolle wie das wachsende Unbehagen angesichts von Big-Data-basierten Methoden, die mit opaken statistischen Verfahren probabilistische Erkenntnis und Prognosen menschlichen Verhaltens generieren.³ Im Vordergrund der gesellschaftlichen Debatte steht dabei in der modernen Informationsgesellschaft die Sorge um die Freiheit und die Privatsphäre des Individuums als eines autonomen Agenten im sozialen Gefüge. Im Hintergrund wird hingegen eine moralphilosophische Problemstellung aufgeworfen, die auf dem Umweg über den ästhetischen Gegenstand auch und gerade die Geisteswissenschaften betrifft: Je treffsicherer individuelles Handeln und Verhalten mathematisch modellierbar werden, desto tiefgreifender wird das seit der Aufklärung propagierte Ideal vom Menschen als einem rationalen, prinzipiell selbstbestimmt und damit selbstverantwortlich handelnden Agenten infrage gestellt. Gerade dieses Selbstbild des Menschen in seinem Verhältnis zu Geschichte und Umwelt aber ist seit der Neuzeit das Kernthema des geisteswissenschaftlichen Gegenstandes ‚Literatur‘.

Die traditionelle philologische Texthermeneutik sieht sich so angesichts der *DH* mit einem methodologischen Paradoxon konfrontiert: Ausgerechnet ihre privilegierte Forschungsdomäne – der ästhetisch-literarische Diskurs – soll nun mit Methoden untersucht werden, die das Geistesleben nicht mehr als existenziellen, kontingenten Phänomenkomplex, sondern als prinzipiell berechen- oder zumindest modellierbaren ‚Output‘ einer multifaktoriellen Funktion konzeptualisieren. Die Skepsis methodisch konservativer Philologen wird dabei weiter befördert durch die programmatische Selbstbeschränkung und Fokussierung der *DH* auf deskriptive oder textdatenbasiert-analytisch bearbeitbare Fragestellungen – oder pointierter gesagt: durch eine Art selbstverordnete hermeneutische Abstinenz der *DH*. Denn Fakt ist: Das Spektrum wie der Grad der methodischen Ausgereiftheit von digitalen Editionsprojekten, stilometrischen Untersuchungen oder diachronen wie systematischen Form-, Struktur- und Themenanalysen mit avancierten NLP-Verfahren haben sich seit der Jahrtausendwende zwar deutlich erweitert. Zu dieser rasanten Entwicklung des Forschungsfeldes *DH* beigetragen haben dabei methodische, technologische, institutionelle und in ganz besonderem Maße auch pädagogisch-curriculare Impulse, kurz: die Gesamtheit der Faktoren, die eine wissenschaftliche Disziplin ausmachen. Der Anschluss an genuin *hermeneutische* Fragestellungen ist hingegen vonseiten der *DH* bislang nur in Ausnahmefällen gesucht worden. Die Praxis digitaler Literaturwissenschaften hat durch die professionelleren, robusteren Verfahren der Datenanalyse und Datenmodellierung vielmehr ein vorwiegend empirisch-analytisches Gepräge gewonnen

³Viktor Mayer-Schönberger/Kenneth Cukier, *Big data. A revolution that will transform how we live, work, and think*, Boston 2013.

und sich damit immer weiter von den spekulativ-exemplarisch verfahrenen Methoden traditioneller Literaturwissenschaften entfernt.⁴

Anders als etwa in den Sprachwissenschaften trifft deshalb das stetig erweiterte Angebot an digitalen Ressourcen, Forschungsinfrastrukturen und immer robuster werdenden analytischen und modellierenden Methoden im texthermeneutischen Bereich auf eine nach wie vor geringe Nachfrage. Digitale Verfahren setzen grundsätzlich eine Konzeptualisierung des Forschungsgegenstandes voraus, die ausschließlich solche Gegenstands- und Prozess-Eigenschaften berücksichtigt, die in Form diskreter Daten erfassbar sind. Gegenstandsfunktionen lassen sich dank dieser Beschränkung damit formal als ein Effekt der Transformation und Verknüpfung von Daten modellieren. Was jedoch jenseits dieser epistemischen Schwelle liegt, ist für *DH*-Verfahren, solange sie primär informationstechnisch orientiert sind, nicht mehr operationalisierbar. Die methodische Konsequenz liegt auf der Hand: Ein Gegenstandsmodell, das beim diskreten Datum ansetzt und von dort aus strikt induktiv und ohne methodische Iteration oder Rückkoppelung *bottom-up* voranschreitet, kann phänomenologisch-hermeneutischen Problemstellungen nicht gerecht werden, die in der Regel ideengeleitet und *top-down*, darüber hinaus aber auch dynamisch und historisch-kontextuell entworfen sind.

Und doch lassen sich auch genuin hermeneutische Operationen des Erschließens, Verstehens und Interpretierens von Gegenständen durchaus mit *DH*-Verfahren unterstützen. Eine zentrale Rolle spielen dabei Verfahren der deklarativen Textauszeichnung und explikativen Textanreicherung, die in den *DH* unter den technischen Begriffen ‚Tagging‘ und ‚Markup‘ verhandelt werden. Das grundsätzliche hermeneutische Potenzial dieser Verfahren lässt sich allerdings erst erkennen, wenn wir die von der Computerlinguistik und der angewandten Informatik übernommene, vom elementaren String her und von vornherein mit Blick auf eine wohlgeordnete Syntax und Grammatik entworfene Konzeption des deklarativen wie des prozessualen Markups überschreiten. Es reicht dabei nicht, zähneknirschend die prinzipielle Interpretationshaltigkeit selbst einer grundlegenden schemakonformen Textauszeichnung mit einem Tag zu konzedieren. Ja, dieses Zugeständnis läuft geradezu auf eine Trivialisierung des Begriffes der ‚Interpretation‘ hinaus: In einem vorgegebenen Kategoriengerüst zwischen zwei

⁴Eine Stichprobe: Für den Zeitraum 2005–2015 werden in GEPRIIS unter dem Suchbegriff ‚Philologie‘ 128 DFG-geförderte Projekte ausgewiesen; nur sieben dieser Projekte (= 5,5 %) weisen zugleich das Suchwort ‚digital‘ auf. (Eine Suchanfrage für ‚Literatur‘ ergibt 2825 Treffer; ‚Literatur & digital‘ ergibt 49 Treffer: eine Quote von 1,7 %.) Aus den online verfügbaren Kurzbeschreibungen geht hervor, dass nur 15 der beschriebenen Projekte einen *DH*-Bezug aufweisen. Von diesen 15 Projekten fallen wiederum zwölf in die Kategorie ‚Digitalisierung/Datenbank‘ und nur drei in die Kategorie ‚Analyse/Annotation‘. Die Suche nach ‚Digital Humanities‘ innerhalb der Literaturwissenschaft ergibt hingegen acht Treffer, wobei jedoch auch hier nur eines dieser Projekte u. a. hermeneutisch arbeitet. Einen ähnlichen Befund ergibt umgekehrt die Sichtung der Tagungsprogramme der regionalen und internationalen *DH*-Fachkonferenzen. Erfasst wurden dabei die Veranstaltungen, die im Zeitraum 2000–2017 von den maßgeblichen Verbänden initiiert wurden, die in der *Alliance of Digital Humanities Organisations (ADHO)* kooperieren.

oder mehr Optionen zu wählen, um ein Phänomen taxonomisch zu deklarieren, ist ein Akt des Sortierens und keiner der Interpretation. Interpretative Akte sind Akte der Kritik, die nicht nur den Gegenstand vermessen und bewerten, sondern immer zugleich reflexiv den Akt des Interpretierens selbst und dessen normative wie methodische Setzungen und Grenzbedingungen thematisieren.

Das Potenzial und die Anschlussfähigkeit von *DH*-Verfahren auch für die hermeneutische Literaturwissenschaft zu erkunden, setzt insofern voraus, die in den *DH* unter dem Terminus ‚Markup‘ etablierten Praktiken der Textauszeichnung mit einem anspruchsvolleren Maßstab zu konfrontieren und sie einmal konsequent vom komplexen, dynamischen Ganzen des Text-Leser-Kontext-Gefüges her zu denken. In dieser Perspektive erweist sich eine Markup-Praxis, die vom lokalen Tagging eines Strings maximal allenfalls bis zur schemakonformen Deklaration eines Gesamttextes reicht, als funktional eingeschränkte, dafür jedoch besser formalisierte technische Variante einer deutlich ambitionierteren ‚weichen‘ Textpraxis, die allen Schriftkulturen seit Jahrhunderten vertraut ist: die der *Annotation von Texten im emphatischen Sinne von hermeneutisch-kommunikativ wie rhetorisch motivierter Informationsanreicherung*. Markup als Annotation in diesem emphatischen Sinne zu konzeptualisieren, bedeutet dabei zugleich, es als ‚methodischen Mittler‘ zu begreifen, der die Anschlussfähigkeit digitaler Verfahren auch und gerade für traditionell-hermeneutisch orientierte Literaturwissenschaftler unter Beweis stellen kann.

2 Annotation als Kommunikation von Lesarten⁵

Annotation als Kulturtechnik ist – methodengeschichtlich gesehen – nicht nur historisch vorgängig, sondern auch variantenreicher als das, was wir gegenwärtig unter den Begriff ‚Markup‘ rechnen – aus der *DH*-Perspektive gesehen mithin eine Praxis *avant la lettre*, die schon lange vor dem *Digital Turn* je nach den spezifischen Anforderungen von Textsorten und Gebrauchskontexten (u. a. religiöse, juridische und ästhetische Exegese) spezifiziert und teilweise auch formalisiert worden ist. Als informationstechnologische Spezifikation stellt Markup insofern nur die jüngste Ausprägung eines übergreifenden Annotationskonzepts dar – eine technisch wie methodisch bedingte Spezifikation, die möglicherweise nicht die letzte bleiben wird.

Die wissenschaftliche Kontextualisierung von Markup als bislang letzte Ausprägung einer lange etablierten Kulturtechnik ist bislang weder ins Bewusstsein der traditionellen Literaturwissenschaften getreten noch umgekehrt von den *DH* als Herausforderung zur Entwicklung eines anspruchsvolleren digitalen Annotationskonzepts angenommen worden. Die im Wesentlichen technologiegetriebene

⁵Ich verwende den Begriff ‚Lesart‘ hier wie nachfolgend im allgemeineren Sinne von ‚Auslegung‘ oder ‚Auffassung‘, also nicht im Sinne des enger gefassten editionsphilologischen Terminus, der den Wortlaut einer Textstelle gemäß eines spezifischen Textzeugens bezeichnet.

Konzeptualisierung von Markup als bloße Anreicherung von Objekten mit deklarativen Metadaten geht vielmehr weiterhin von einem erheblich engeren, instrumentell gefassten Begriff der ‚Anreicherung‘ aus. Als Markup in diesem engeren Sinne gelten systematisch geregelte Beschreibungen des Gesamttextes wie lokale Textauszeichnungen auf Wort- und Zeichenebene, die unter Rückgriff auf Standards, Taxonomien und Objektontologien und in der Regel für definierte Anschlussoperationen erstellt werden. Für die Objektbeschreibung auf elementarer Ebene sind prototypisch die Verfahren der linguistischen Basisannotation wie z. B. *POS Tagging*; für die systematische wie historische Einordnung des Gesamtdokumentes sind es etwa die Hinzufügung von Metadaten in einem TEI-Header. Generische Schemata wie GML, SGML, HTML bis hin zu TEI und den diversen heute gebräuchlichen XML-basierten domänenspezifischen Markup-Sprachen setzen dabei konzeptionell ein zwar abstraktes, aber klar definiertes Objektmodell voraus – sei es das grammatische Modell einer natürlichen Sprache, das jedes *Part-of-Speech Tagging* organisiert, sei es das einer *Ordered-Hierarchy-of-Content-Objects-Sicht (OHCO)* auf Textdokumente, das bei der TEI-konformen Textauszeichnung im Hintergrund steht. In der konkreten Annotationspraxis wird dieses abstrakte Objektmodell dann mithilfe einer geregelten Syntax am konkreten Objekt *bottom-up* instanziiert. Das Objektmodell selbst steht dabei prinzipiell nicht mehr zur Debatte – es sei denn, man stößt bei der Anwendung auf grundsätzliche Probleme, wie die 1999 mit einer Kontroverse zwischen Renear, McGann und Buzzetti einsetzende Debatte um die Unvereinbarkeit des *OHCO*-Modells mit dem Faktum der sog. *Overlapping Hierarchies* gezeigt hat.⁶

Coombs u. a.⁷ zählen zu den ersten, die die auf GML aufbauenden konkreten technischen Umsetzungen der taxonomie- und modellgetriebenen Markup-Praxis systematisch und funktional differenziert haben. Sie unterscheiden zwischen den Formen *Punctuational*, *Presentational*, *Procedural*, *Descriptive*, *Referential* und *Meta-Markup*. Als für den menschlichen Leser besonders relevante Variante heben die Autoren dabei das sog. *Descriptive Markup* hervor.⁸ Ähnlich weist 25 Jahre später

⁶Siehe hierzu Susan Hockey/Jerome J.McGann/Allen Renear, „What is text? A debate on the philosophical and epistemological nature of text in the light of humanities computing research“, in: *1999 ACH-ALLC Conference Program*, <http://www2.iath.virginia.edu/ach-allc.99/proceedings/hockey-renear2.html> (letzter Aufruf 7.5.2018). Mit diesem Beitrag rückte die Thematik auf der ACH-ALLC-Konferenz 1999 erstmals ins Zentrum der Aufmerksamkeit. Dino Buzzetti trug schon auf der gleichen Konferenz mit dem Vortrag „Text Representation and Textual Models“ explizit zur Debatte bei, in: *1999 ACH-ALLC Conference Program*, <http://www2.iath.virginia.edu/ach-allc.99/proceedings/buzzetti.html> (letzter Aufruf 7.5.2018; im Beitrag „Digital Representation and the Text Model“, in: *New Literary History* 33 (2002), 61–88, legte Buzzetti seine Argumentation dann umfassender dar.

⁷James H. Coombs/Allen H. Renear/Steven J. DeRose, „Markup systems and the future of scholarly text processing“, in: *Communications of the ACM* 30 (1987), 933–947.

⁸Siehe jedoch Allen Renear, „The descriptive/procedural distinction is flawed“, in: *Markup Languages* 2/4 (2000), 411–420, der später feststellt: „the descriptive/procedural distinction is flawed“.

auch Nyhan darauf hin, dass die nicht bereits auf spezifischere Anschlussoperationen der maschinellen Transformation oder Präsentation eines Dokumentes zugeschnittene Variante einer beschreibenden Annotation flexibel und auf verschiedenste Textsorten, Genres und Medien übertragbar ist: „Descriptive markup can be applied to any kind or genre of text; indeed, any information that can be consistently represented using a symbol of some kind and then digitized can be marked up“.⁹

Nochmals zu betonen ist allerdings, dass das ‚beschreibende‘ Markup, von dem hier die Rede ist, *de facto* nicht mehr als ein kategorial sortierendes, allenfalls taxonomisch evaluierendes Verfahren ist. Das Beschreibungsschema, die Beschreibungskategorien sowie der Typ und das Spektrum der zulässigen Werte, die dem jeweils markierten Objektbereich – einem Morphem, Wort, Satz, Absatz etc. – zugewiesen werden können, sind geregelt – und nicht *ad hoc* erweiter- oder veränderbar. Deklariert wird vielmehr anhand einer vorgegebenen Systematik und Metrik, die zudem gegenüber etwaigen Forschungsfragen agnostisch ist und dies auch sein muss, um überhaupt generalisierbar sein zu können. Ein Textdokument in TEI zu beschreiben oder eine verschriftlichte sprachliche Äußerung zu parsen und mit *POS Tags* zu versehen, läuft so, genau besehen, darauf hinaus, das jeweils selektierte Textelement in eine vorgegebene Tabelle einzusortieren und einen (oder mehrere) Werte aus einem finiten alphanumerischen Wertebereich zuzuordnen.

Zu erinnern ist hier an die von Gadamer getroffene Unterscheidung zwischen dem Verfahren des ‚Beschreibens‘, das die empirischen Wissenschaften anwenden, und jenem des ‚Erklärens‘, das für die Geisteswissenschaften zentral ist. Wittgenstein liefert mit einem Satz in seinen *Philosophischen Untersuchungen* ein gutes Beispiel, was geschieht, wenn man diese Distinktion zur programmatischen Norm erhebt. Dort heißt es zunächst: „Alle Erklärung muß fort, und nur Beschreibung an ihre Stelle treten“;¹⁰ Deskription soll in den Wissenschaften herrschen, nicht Interpretation.¹¹ Aber im Nachsatz räumt Wittgenstein dann ein: „Und diese Beschreibung empfängt ihr Licht, d. i. ihren Zweck, von den philosophischen Problemen“.

Auch wenn man im Grundsatz eine klare und bewusst praktizierte Trennung der beiden Operationen des Beschreibens und des Erklärens auch von der Philologie fordern sollte, muss man zur Kenntnis nehmen, dass in der Praxis der wissenschaftlichen Texthermeneutik die zweckorientierte, philosophisch motivierte ‚Beschreibung‘ des Gegenstandes im emphatischen Sinne immer auf eine ‚Erklärung‘ hinauslaufen wird – mithin auf eine Interpretation, die als Verfahren der Gegenstandserschließung nicht nur auf den Gegenstand selbst gerichtet ist, sondern zugleich reflexiv und dynamisch eben dieses Verfahren der

⁹Julianne Nyhan, „Text encoding and scholarly digital editions“, in: Dies./Claire Warwick/Melissa Terras, *Digital Humanities in Practice. Facet Publishing*, London 2012, 117–138, hier: 123.

¹⁰Ludwig Wittgenstein, *Philosophische Untersuchungen*, Frankfurt a. M. 1975 [1953], 78 f.

¹¹Zur Anwendung Wittgensteins auf die Problematik der Literaturtheorie siehe u. a. Henry Staten, „Wittgenstein and the Intricate Evasions of ‚Is‘“, in: *New Literary History* 19/2 (1988), 281–300.

Erschließung selbst thematisiert. Diese unkalkulierbaren Rückkoppelungseffekte sind nun jedoch genau das, was ein standardisiertes, auf Disambiguierung verpflichtetes Markup auszuschalten versucht.

Der methodologische Reduktionismus, der dem deklarativen ebenso wie dem prozeduralen oder repräsentationellen Markup insofern inhärent ist, hat in den *DH* mehrfach die Forderung nach der Entwicklung eines interpretativen oder hermeneutischen Markupkonzepts ausgelöst. Piez fordert in diesem Sinne prononciert ein „markup that is deliberately interpretive“; diese Art Markup sei dabei „not limited to describing aspects or features of a text that can be formally defined and objectively verified. Instead, it is devoted to recording a scholar’s or analyst’s observations and conjectures in an open-ended way.“¹² Wie schon Caton hervorgehoben hat, ist insofern die Frage der Wahl eines Markupschemas in der hermeneutischen Perspektive eine nachgeordnete; ausschlaggebend ist vielmehr das Textkonzept an sich:

When OHCO-1 encourages encoders to see a written text as a thing, they stay above the content and only drop down to engage with the text as message to identify the occasional editorial object whose nature is not obvious from its appearance. But when encoders see the written text as a communicative act, they must participate in the act: take on the role of hearer, attend to what the text says, and identify the speaker’s intentions not just from the words’ semantics but also from the attitudes conveyed. Metaphorically, encoders must be down at what would be the lowest level of an OHCO tree, completely immersed in the #PCDATA, because content generates interactional encoding far more than any content object. As its practitioners well know, all encoding interprets, all encoding mediates. There is no ‚pure‘ reading experience to sully. We don’t carry messages, we reproduce them – a very different kind of involvement. We are not neutral; by encoding a written text we become part of the communicative act it represents.¹³

Das hier skizzierte Modell eines *Interactional Encodings* zieht einen konzeptionell erheblich weiter gefassten Begriff von ‚Markup‘ nach sich, als er mit einer am Ideal des perfekten *Inter Annotator Agreement* ausgerichteten Auszeichnungspraxis einhergeht – einer Praxis, die aus der Linguistik importiert wurde und damit in der Tradition der Bemühungen um eine Verwissenschaftlichung der Literaturwissenschaften steht, die schon zu Beginn des 20. Jahrhunderts mit dem Russischen Formalismus einsetzte.¹⁴ In den methodengeschichtlich bislang eher desinteressierten *DH* unserer Tage wird dieses Ideal des non-ambigen Enkodierens und Auszeichnens allerdings weniger programmatisch als vielmehr pragmatisch legitimiert: *Gold Standards* spielen für die Automatisierung und Validierung von Markup-Routinen mit Verfahren des ML bekanntlich eine zentrale Rolle.

¹²Wendell Piez, *Towards Hermeneutic Markup. An architectural outline* (2010), <http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/html/ab-743.html> (letzter Aufruf 9.11.2017).

¹³Paul Caton, „Markup’s Current Imbalance“, in: *Markup Languages 3/1* (2000), 1–13.

¹⁴Ein Faktum, dem interessanterweise Rumelhart als Pionier der textbezogenen AI-Forschung der 1970er Jahre bereits implizit Rechnung getragen hat, indem er wiederholt auf Propps *Morphologie des Zaubermärchens* zurückgriff – wenn auch in einer Lesart, die bezeichnenderweise das Konzept der Morphologie, das Propp mit einem seinem Buch vorangestellten Goethe-Motto geistesgeschichtlich kontextualisiert hatte, auf einen generativen Algorithmus reduziert.

Diese in der Computerlinguistik, der NLP-Forschung und der digitalen Texteditorik verbreitete Konzeption einer am Ideal der Eindeutigkeit orientierten Textauszeichnung wird selbst in jüngeren *DH*-Handbüchern und Einführungen weitgehend alternativlos präsentiert.¹⁵ In dieser Verengung der Perspektive drückt sich eine methodische Problematik aus, die van Zundert jüngst als die der – bislang ungenügend berücksichtigten – *Computationality of Hermeneutics* umreißt.¹⁶ Van Zundert verbindet seinen historischen Abriss der in den *DH* zur Zeit nur noch am Rande geführten Forschungsdebatte zur Frage des methodologisch-philosophischen Verhältnisse zwischen den hermeneutischen Wissenschaften und der Informatik entsprechend mit einem Plädoyer, das hermeneutische Anliegen nicht länger nur „after the algorithmic fact“ zu thematisieren. Gefordert sei vielmehr, hermeneutische Desiderata bereits auf der Ebene der Konzepte, Codes und Modelle zu verhandeln, die für die digitale Repräsentation und Analyse von lebensweltlichen Phänomenen und ästhetischen Artefakten grundlegend sind.

Was aber wäre ein solches der Texthermeneutik wie der Informatik gemeinsames und damit für die *DH* kennzeichnendes hermeneutisches Anliegen ‚before the algorithmic fact‘? Der allen Varianten metasprachlicher Äußerung gemeinsame rhetorische Gestus ist der einer Kommunikation einer (oder mehrerer) Lesart(en) eines Objekttextes – und sei es nur in der Form eines -Tags. Eine typographische Lesarten-Kommunikation mag zwar vordergründig nur noch die an eine Maschine gerichtete Instruktion sein, eine Zeichenfolge zu fetten. Aber selbst diese vermeintlich nur prozessuale Codierung ist ihrerseits Mittel zum Zweck: Auch sie richtet sich in letzter Instanz wieder an einen menschlichen Leser und ist ursprünglich diskurspragmatisch legitimiert.

Die hier vorgeschlagene diskurspragmatische Konzeptualisierung von Annotation als Kommunikation von Lesarten lässt zwei Grenzbedingungen erkennen: Annotation ist einerseits notwendig eine Form von Metatext, eine Rede über die Rede – wo sie sich vom Bezugstext ablöst und autonom wird, nimmt sie ihrerseits objektsprachlichen Charakter an und wird zur selbstständigen Sachverhaltsbehauptung. Andererseits verliert Annotation ihren kennzeichnenden Kommunikationscharakter aber auch dort, wo Metasprache zur reinen Privatsprache wird. Ein Beispiel dafür liefert das jüngst von McGann vertretene Konzept des sog. ‚topologischen‘ Markups.¹⁷ McGann kritisiert das TEI-Modell vor allen Dingen deshalb, weil es die autopoetische Funktion von ästhetischen Texten ausklammere zugunsten einer allopoetischen Funktion, die Bedeutung nur noch als

¹⁵ Siehe Andrea Rapp, „Manuelle und automatische Annotation“, in: Fotis Jannidis/Hubertus Kohle/Malte Rehbein (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart 2017, 253–267, hier: 255 f.; Allen Renear, „Text Encoding“, in: Susan Schreibman/Ray Siemens/John Unsworth (Hg.), *Companion to Digital Humanities*, Oxford 2004.

¹⁶ Joris J. van Zundert, „Screwmenetics and Hermenumericals. The Computationality of Hermeneutics“, in: Susan Schreibman/Ray Siemens/John Unsworth (Hg.), *A New Companion to Digital Humanities*, Chichester, West Sussex, UK, u. a. 2016.

¹⁷ Jerome McGann, Jerome, „Marking texts in many dimensions.“, in: *A new republic of letters: Memory and scholarship in the age of digital reproduction*, Cambridge, MA, 2014, 90–112.

ein dem Text externes Output denkt, nicht aber als das dynamische Produkt eines Text-Leser-Komplexes *in actu*. Sein Gegenentwurf geht deshalb von einer als *Patacriticism* bezeichneten ‚theory of subjective interpretation‘ aus – und läuft in seiner Radikalität dann auf eine Annotation zum Zwecke der Autokommunikation des Annotators hinaus: eine Privatsprache, die keine Diskurspartner mehr sucht und den annotierten Text als Impulsgeber eines Selbstgesprächs begreift.¹⁸

Die von mir vorgeschlagene diskurspragmatische Bestimmung von Annotation als Kommunikation von Lesarten positioniert die Annotation hingegen als eine Art hermeneutischen Mediator, der in der Mitte zwischen dem Objekttext und seinem Gebrauchskontext funktional wird. Dies entspricht der von Boot unter dem Begriff *Mesotext* diskutierten epistemischen Funktion von Annotationen als einem Mittler zwischen dem (Objekt-)Text und dem medial eigenständigen Metatext.¹⁹ Wichtig ist dabei insbesondere Boots Stufung *Textdaten – Mesodaten – Mesotext*, d. h. der Hinweis darauf, dass auch Annotationen immer bereits unter zwei Voraussetzungen generiert und verarbeitet werden: erstens einer der Annotationspraxis inhärenten Texttheorie, die eine deskriptive bzw. explikative Modellierung des Objekts ‚Text‘ überhaupt erst ermöglicht, und zweitens einer Ausrichtung auf ein diskursives ‚Narrativ‘, das die per Annotation formal erfassten einzelnen Beobachtungen, Erläuterungen und lokalen Interpretationen zum Objekttext sinnhaft integriert, mithin aus den isolierten Mesodaten überhaupt erst Information und damit einen Mesotext macht.

Für eine hermeneutisch ambitionierte *DH*-Praxis wäre insofern ein Begriff von ‚Markup‘ zu fordern, der nicht mehr vorwiegend technologisch, sondern vielmehr diskurspragmatisch gedacht ist. Wenn wir Markup als *DH*-spezifische Ausprägung von ‚Annotation‘ begreifen wollen, dann muss das Markup-Konzept auch Praktiken jenseits der bloßen Codierung oder Deklaration von Objektdaten einschließen – denn nur dann können auch die noch nicht auf spezifische informationstechnische Anschlussoperationen hin orientierten philologischen Verfahren der Anreicherung von textuellen Primärdaten mit Zusatzinformationen berücksichtigt werden, die einer hermeneutisch motivierten Kommunikation von Lesarten dienen. Ob und vor dem Hintergrund welcher Modelle diese Lesarten dann ihrerseits evaluiert werden, bleibt dabei eine Frage des spezifischen Anwendungsfalls und seiner Erfordernisse. Widerspruchsfreiheit und Konsistenz sind insofern keine intrinsischen, formalen Kriterien von Annotationen, die als solche normative Gültigkeit beanspruchen könnten – es sind dies vielmehr aus dem Anwendungskontext heraus definierte Anforderungen. Entsprechend wertneutral wäre deshalb auch das Konzept des ‚Markups‘ zu fassen, dessen Maßstab nicht eine abstrakte logische Spezifikation sein kann, sondern nur das aus einem Anwendungskontext heraus formulierte und vor dem Hintergrund einer jeweiligen Texttheorie spezifizierte Markupschema.

¹⁸Das McGanns Artikel zur Illustration beigefügte Beispiel einer autopoetisch inspirierten Annotation und Lektüre liest sich dann auch wie ein *Remake* von Roland Barthes' *S/Z*.

¹⁹Peter Boot, *Mesotext. Digitised Emblems, Modelled Annotations and Humanities Scholarship*, Utrecht 2009.

3 Funktionen und Formen von Annotationen

Nahezu sämtliche Varianten philologisch fundierter Lesartenhinweise an den Leser können heute als Verfahren digital emuliert und als Mitteilung digital codiert werden. Als wissenschaftliche Operation ist die Textannotation dabei funktional für drei der sog. *Scholarly Primitives* nach Unsworth: Objektkonstitution, Objektbeschreibung und Objektannotation.²⁰ Im engeren Anwendungsbereich der Literaturwissenschaften lassen sich analog drei Komplexitätsebenen unterscheiden:

- *Textauszeichnung*: Auszeichnung von Phänomenen, die keine inhärente oder robust konventionalisierte Semantik aufweisen. Dazu gehören sowohl rein formale Phänomene wie etwa Layout und Typographie als auch strukturell-grammatische Phänomene. Vor der Auszeichnung liegt hier i. d. R. ein eindeutiges Auszeichnungsschema vor (linguistische Taxonomien, TEI etc.).
- *Textannotation*: Auszeichnung lokaler semantischer Phänomene. Diese Variante der Auszeichnung setzt bereits das Verstehen des semantischen Gehalts einer umfangreicheren Textstelle (und damit meistens deren Interpretation im weiteren Sinne²¹) voraus. Diese auf der mittleren semantischen Komplexitätsebene ansetzende Auszeichnung lässt sich im Anschluss an Piez²² als ‚hermeneutische Annotation‘ bezeichnen. Auch wenn in der Praxis der Textannotation qua Verwendung literaturwissenschaftlicher Fachterminologie oftmals eine Art implizite Taxonomie vorausgesetzt ist, wird diese deutlich weniger deterministisch gehandhabt als bei der schemabasierten Textauszeichnung.
- *Textkommentar*: Auszeichnung dezidiert literarischer Bedeutungsphänomene, die sich in der Regel nur durch eine holistische Interpretation des Gesamttextes erschließen. Dazu gehören beispielsweise die allegorische oder parabelhafte Auslegung von Textbedeutungen sowie nur interpretativ lösbare Fragen nach ‚Wahrheit‘ im Kontext fiktionaler Repräsentationen. Für die Auszeichnung derartiger Phänomene ist eine komplexe Interpretationstheorie für literarische Texte erforderlich und, sofern interpretationstheoretisch abgesichert, darüber hinaus auch die Referenz auf kontextuelle (d. h. textexterne) Informationen. Textkommentare sind insofern noch weniger deterministisch und deshalb kaum sinnvoll digital modellierbar.

Diese systematische Differenzierung ist entlang der Achse zunehmender semantischer und damit auch hermeneutischer Komplexität von Annotationen entworfen.

²⁰John Unsworth, *Scholarly Primitives. What methods do humanities researchers have in common, and how might our tools reflect this?*, London 2000.

²¹Siehe Janina Jacke, „Is There a Context-Free Way of Understanding Texts? The Case of Structuralist Narratology“, in: *Journal of Literary Theory* 8/1 (2014), 118–139, hier: 130 f.

²²Piez (Anm. 13).

Eine alternative Systematik schlagen Bauer/Zirker vor,²³ die teils unter dem Gesichtspunkt des Gegenstandsbereiches (Typen A und B) und teils unter dem der explikativen Relation einer Annotation zum Text (C, D, E) fünf Annotationstypen unterscheiden:

- A linguistic
- B formal
- C intratextual
- D contextual
- E interpretive

Beide Systematisierungen ermöglichen es, das gesamte Spektrum an ‚Lesarten-hinweisen‘ zu erfassen, die uns konkret in Form von Hervorhebungen, Unterstreichungen, in den Text direkt eingefügten Anmerkungen, Kommentaren in der Marginalienspalte usw. begegnen.

In einer dritten, auf den ersten Blick rein technischen Perspektive lassen sich Annotationstypen schließlich auch nach der medialen bzw. technischen Distanz einer Annotation zum jeweiligen Referenzbereich unterscheiden. Die im Printmedium geläufigen Formen der typographischen Hervorhebung wie z. B. Fettungen oder Unterstreichungen nutzen dabei das objektsprachliche Zeichen selbst als Träger der metasprachlichen Annotation. In den heute geläufigen SGML-basierten digitalen Varianten treten die objekt- und die metasprachlichen Dimensionen hingegen bereits auf dieser grundlegenden Ebene auseinander.²⁴ Das ist jedoch weder zwingend, noch wird dieses technische Detail in der Regel vom Normalleser eines digitalen Dokuments überhaupt bewusst verarbeitet. In der Perspektive des technisch interessierten Lesers hingegen läuft die in der digitalen Zeichenkette eingebettete Deklaration mittels *Inline Tag* auf eine doppelte Emulation hinaus: Erstens sichert das *Inline Tagging* die Eindeutigkeit der Objektreferenz auf eine quasi-analoge Weise durch die unmittelbare räumliche Nachbarschaft von Objekt- und Metazeichen; zweitens emuliert der digitale Text in den meisten Arbeitsumgebungen dann spätestens auf dem Bildschirm wieder die Verschmelzung der typographischen Annotation mit dem objektsprachlichen Zeichen.

Auf den zweiten Blick allerdings zeigt sich, dass die Gradierung von Annotationen nach ihrer medialen Distanz zum Objektzeichen bereits in der prädigitalen Textpraxis eine funktionale Semantisierung erfahren hat: Je weiter das

²³ Matthias Bauer/Angelika Zirker, „Whipping Boys Explained. Literary Annotation and Digital Humanities“, in: Kenneth M. Price/Ray Siemens (Hg.), *Literary Studies in the Digital Age. An Evolving Anthology*, New York 2013, <https://dlsanthology.mla.hcommons.org/> (letzter Aufruf 3.5.2018).

²⁴ Was durchaus keine technologische Notwendigkeit ist, sondern eine Konsequenz von SGML: Im Prinzip ließen sich mit einem Hexadezimal-Code sämtliche typografischen Varianten eines Zeichens auch einzeln codieren und dennoch zugleich auf ähnliche Weise als funktional ‚familienähnlich‘ behandeln, wie wir dies bei der digitalen Codierung von Farbwerten tun.

objektsprachliche Zeichen und die metasprachliche Annotation auseinandertreten, desto eher wird ein kompetenter Leser die Annotation als interpretationshaltigen Mesotext *sensu* Boot bewerten, der einem eigenständigen Metatext zuarbeitet. Im traditionellen Schrift- und Printmedium beginnt dieses Auseinandertreten mit der Interlinearannotation und reicht über die Formen der Glosse und der Anmerkung in der Marginalienspalte bis hin zur komplexen, kommentierenden Fuß- oder Endnote und möglicherweise gar noch bis zum Apparat. Die mediale und räumliche Distanz zwischen Objekttext und Annotat ist so zum konventionalisierten Indikator diskurspragmatischer bzw. rhetorischer Funktionalität geworden.

4 **Markup als Annotation jenseits des *Document Paradigm***

In der Perspektive der *DH* und mit Blick auf Markup als digitale Variante von Annotation stellt sich an dieser Stelle eine Grundsatzfrage: Alle Markuptechniken, die in der Tradition von SGML stehen, setzen konzeptionell auf die durchgehende Separierung von Objekttext und Metatext. Je deutlicher diese Trennung im Interface und im Workflow einer Markupumgebung ausgeprägt sind, umso sperriger wirkt diese daher für den traditionell arbeitenden Forscher, der Annotation als eine weitgehend unstrukturierte, intuitiv handhabbare Praxis der Anreicherung und Zwischenspeicherung von Anmerkungen, Erläuterungen, Kommentaren und Verweisen schätzt.

Die Suggestivität, Flexibilität und Reichhaltigkeit dieser ‚händischen‘ Annotationspraxis lässt sich zwar mit einigem Aufwand auch digital simulieren, wie John Bradley mit der Entwicklung der Arbeitsumgebung *Pliny* gezeigt hat.²⁵ Aber warum überhaupt an dem prädigitalen Muster festhalten und es zu emulieren suchen? Macht nicht gerade die Entkoppelung des digitalen Markups von der konventionalisierten Semantik ‚händischer‘ Annotation, die eine Anmerkung als umso selbstevidenter und ‚objektiver‘ suggeriert, je geringer die räumliche Distanz zwischen Objekttext und Annotat ist, einen *epistemischen Mehrwert* digitaler Annotationen aus?²⁶ Ist das, was auf den ersten Blick einer informationstechnischen Gegebenheit geschuldet zu sein scheint, nicht eigentlich Ausdruck einer prinzipiell anderen Form von Wissensorganisation und die logische Konsequenz eines neuen Textmodells? Welche neuen Formen, Funktionen und

²⁵ John Bradley, „Thinking about interpretation. Pliny and scholarship in the humanities“, in: *Literary and Linguistic Computing* 23 (2008), 263–279.

²⁶ Interessant ist hier die Frage nach dem funktionalen wie rhetorischen Status eines *inline* codierten Hyperlinks: Wird die Referenz vom Objekttext zum Zieltext damit epistemisch objektiviert und vom Leser als Quasi-Merkmal des Ursprungstextes verarbeitet, weil der Link auf dem Bildschirm mit seinen Trägerzeichen verschmolzen ist – oder wird umgekehrt die Verweisfunktion des Links gerade bewusst als solche wahrgenommen, weil sie erst durch eine Interaktion aktiviert werden muss, die das Trägerzeichen als Link interpretiert?

Praktiken literaturwissenschaftlicher, hermeneutisch ambitionierter Annotation könnten mit diesem Ansatz entstehen?

Eine der zentralen Debatten innerhalb der Markup-Community ist in dieser methodologischen Hinsicht die Kontroverse um die Verwendung von *Inline* vs. *Stand-off Markup*. *Inline Markup* in Form von in den Objekttext eingebetteten Tags ist bei der sequenziellen Verarbeitung von Dokumenten leichter handhabbar, konzeptionell den einfachsten Formen des analogen Verfahrens ähnlich und bietet zudem eine Reihe von Vorteilen bei der Datenverarbeitung (Performanz von Search- und Retrieval-Algorithmen, Stabilität der Referenz u. a. m.). Die Alternative des explizit referenzierenden *Stand-off Markups* bzw. *External Stand-off Markups*, in der das digitale Annotat konsequent vom Bezugstext abgelöst und separat gespeichert wird, kann hingegen besser komplexere Bezüge abbilden, die nicht mehr an die formale Strukturierung des Bezugstextes als lineare Zeichenkette gebunden sind.

Ein wesentliches Argument für die Verwendung von *Stand-off-Markup-Verfahren* ist bekanntlich die Kritik am hierarchischen Textmodell der *Ordered Hierarchy of Content Objects (OHCO)* gewesen, wie es etwa dem TEI-Schema zugrunde liegt. Das OHCO-Modell von ‚Text‘ ist entlang einer hierarchisch geordneten Systematik nach dem Muster ‚Buch (enthält Kapitel (enthält Absatz (enthält Satz (enthält Wort)))‘ entworfen; es lässt sich damit problemlos und performant mit einem *Inline* Markup abbilden, dessen Syntax die entsprechende hierarchische Staffelung der Tags garantiert. Die sog. ‚OHCO-Debatte‘ setzte vor diesem Hintergrund zunächst bei einem technischen Detail und damit *bottom-up* bei einem Praxisproblem an: Wie lässt sich eine sog. *Overlapping Hierarchy* codieren? Ein solcher *Overlap* liegt zum Beispiel bei einem lyrischen Enjambement vor, in dem eine Phrase die Versgrenze überspringt. *Inline* Markup stößt hier an eine Grenze, denn der schließende Tag der hierarchisch untergeordneten syntaktischen Einheit ‚Phrase‘ liegt jenseits des schließenden Tags der übergeordneten Einheit ‚Vers‘.²⁷

Stand-off Markup ist eine von mehreren Möglichkeiten, dieses Detailproblem der Markuppraxis technisch zu lösen.²⁸ Aber das ist eben nur die technische,

²⁷ Zur Problematik der *Overlapping Hierarchies* insbesondere im Kontext von TEI siehe Elena Pierazzo, „Textual Scholarship and Text Encoding“, in: Susan Schreibman/Ray Siemens/John Unsworth (Hg.), *A New Companion to Digital Humanities*, Chichester/West Sussex u. a. 2016, 307–321, hier: 316–319, sowie Andreas Witt, „Multiple hierarchies. New aspects of an old solution“, in: *Proceedings of Extreme Markup Languages (2004)*, <http://conferences.idealliance.org/extreme/html/2004/Witt01/EML2004Witt01.html> (letzter Aufruf 8.1.2017).

²⁸ Eine andere Möglichkeit ist z. B. die Codierung in sog. *Goddags (General Ordered-Descendant Directed Acyclic Graphs)*. Siehe hierzu Yves Marcoux/Michael Sperberg-McQueen/Claus Huitfeldt, „Modeling overlapping structures“, in: *Balisage. The Markup Conference (2013)*, <https://www.balisage.net/Proceedings/vol10/html/Marcoux01/BalisageVol10-Marcoux01.html> (letzter Aufruf 26.4.2018).

problemorientierte Perspektive. Die *konzeptionelle* Relevanz von *Stand-off Markup* für die Praxis digitaler Textwissenschaften ist grundsätzlicherer Natur: Denn überlappende Hierarchien und diskontinuierlich über das Ganze ausgeprägte Phänomene sind nicht nur ein formales Binnenmerkmal des literaturwissenschaftlichen Gegenstandes ‚Text‘ – sie kennzeichnen vielmehr auch den pragmatischen Kontext und die methodischen Praktiken, innerhalb dessen Texte als Objekte der Literaturwissenschaften seit dem *Digital Turn* thematisch werden. Die traditionelle philologische Praxis ließ sich noch denken unter Bezugnahme auf eine hierarchisch-substanzorientierte Vorstellung von Texten (primären Gegenständen und sekundären Informationsquellen), Bibliotheken (Institutionen der Quellen- und Wissensverwaltung) und eine klare Distinktion der Rollen von Autoren (intentional agierenden Produzenten) und Lesern (Adressaten und Interpreten der verschriftlichten Rede). Für eine Philologie, die das Erkenntnispotenzial digitaler Medien, Modelle und Verfahren konzeptionell für sich nutzen will, greift diese aus der Ära des Printmediums stammende Modellierung von Texten und Gebrauchskontexten, die sich u. a. in dem deskriptiv-hierarchischen Markupmodell niederschlägt, jedoch zu kurz.

Die digitale Praxis literaturwissenschaftlicher Textthermeneutik sollte diesen Emulationsmodus deshalb konsequent hinter sich lassen und nicht länger von einer essenzialistischen Vorstellung von Text und traditionellen Rechte-Rollen-Schemata ausgehen, die in der allgemeinen literaturtheoretischen Debatte ohnehin spätestens mit dem Poststrukturalismus infrage gestellt worden sind. Eine Praxis der „literary annotation in the digital age“;²⁹ die sich nicht auf die leichter digital zu operationalisierenden Fragestellungen formal orientierter Arbeitsfelder wie etwa Narratologie oder Editorik beschränkt, wird deshalb vermutlich ein insgesamt eher ereignisorientiertes denn ein substanz- und rollenorientiertes Textmodell entwickeln müssen. Für die digitale Annotation bedeutet dies, den Objekttext wie seine Annotationen mal als Knotenpunkte, mal als Kanten in einem n-dimensionalen, dynamisch rekonfigurierbaren Netzwerk zu denken. Einer der Akteure in diesem Netzwerk ist dabei der Textinterpret: Je nach Erkenntnisinteresse und methodischem Ansatz definiert, systematisiert und exploriert er als hermeneutischer Agent im Zuge der digitalen Annotation die ‚Kanten‘ zu und von den Texten und ihren Annotaten sowie deren mögliche Clusterungen. Ein – wenn nicht gar der zentrale – Vorteil digitaler Repräsentationen von textuellen Objekten und Bezügen gegenüber analogen Vorläufern und Praktiken ist dabei die erheblich einfachere Aggregation, Rekonfiguration und Analyse der textuellen Objekte und Annotate, die jeweils in den Blick genommen werden können. Ein zweiter Vorteil ist die Möglichkeit zum ‚Re-Run‘ einer Forschungsfrage, die digitale Verfahren dank formaler Parametrisierung und computergestützter Abarbeitung repetitiver Routinen eröffnen.

²⁹ Bauer/Zirker (Anm. 21).

Tab. 1 Prototypische Ausprägungen von digitalen Annotationen als Praktiken der Textmodellierung

	konzeptionelles Modell	logisches Modell
diskursives digitales Annotieren	hermeneutische Annotation	explorativer Lesartenhinweis
deklaratives digitales Annotieren	taxonomische Annotation	formalisierter Lesartenhinweis

Für den engeren Bezugsbereich einer digitalen Texthermeneutik bedeutet dies, dass die einzelne Textinterpretation nicht mehr zwingend als geschlossenes Ganzes gedacht werden muss, sondern dass sie auch als eine unter n möglichen Konfigurationen von Interpretations- und Deklarationsentscheidungen konzeptualisiert werden kann. *Stand-off* Markup unterstützt dieses Verfahrensmodell, weil es konzeptionell nicht mehr ein dokumentenzentriertes Verfahren ist, sondern ein datenbankzentriertes, wie Schloen/Schloen hervorheben:

[...] stand-off markup deviates so much from the original markup metaphor that it no longer belongs within the document paradigm at all and is best implemented within the database paradigm. Stand-off markup involves the digital representation of multiple readings of a text by means of separate data objects, one for each reading, with a system of pointers that explicitly connect the various readings to the text's components. But this amounts to a database solution to the problem. The best way to implement this solution is to abandon the use of a single long character sequence to represent a scholarly text — the document approach — in order to take advantage of the atomized data models and querying languages characteristic of database systems.³⁰

Diese digitale Repräsentation von Annotationen als „multiple readings of a text by means of separate data objects“ setzt allerdings noch mehr voraus als nur den Wechsel vom dokumentenzentrierten zum datenbankzentrierten Modell: Sie bedeutet auch, die Praxis der digitalen Annotation insgesamt als eine des *Data Modeling* zu betrachten. In Anlehnung an Flanders/Jannidis³¹ ließe sich dann die Praxis der Datenmodellierung vorgeschlagene Unterscheidung *Conceptual Model* vs. *Logical Model* und *Curation-Driven Modelers* vs. *Research-Driven Modelers* in abgewandelter Form zu einer Matrix erweitern, mit der sich vier prototypische Ausprägungen von digitalen Annotationen als Praktiken der Textmodellierung unterscheiden lassen (s. Tab. 1).

³⁰David Schloen/Sandra Schloen, „Beyond Gutenberg. Transcending the Document Paradigm in Digital Humanities“, in: *DHQ* 8/4 (2014), <http://www.digitalhumanities.org/dhq/vol/8/4/000196/000196.html> (letzter Aufruf 3.5.2018).

³¹Julia Flanders/Fotis Jannidis, „Data Modeling“, in: Susan Schreibman/Ray Siemens/John Unsworth (Hg.), *A New Companion to Digital Humanities*, Chichester, West Sussex, UK, ²2016, 229–237.

Diskursiv orientiertes, hermeneutisch funktionales Annotieren als Exploration und Verschriftlichung individueller ‚Lesarten‘ eines Textes wird einen konzeptionellen Nutzen aus der Digitalisierung des Verfahrens in vielen Fällen zunächst nur mittelbar ziehen können: Je kontextabhängiger die Interpretation einer Textstelle ist, desto schwieriger ist die Formalisierung des Prozesses. Auch für die hermeneutisch ambitionierte Annotationspraxis mag das automatisierte *Pre-Processing* und ggf. auch das automatische Taggen eines Textes zwar durchaus heuristischen Wert haben – die eigentliche Aufgabe der Textinterpretation ist damit jedoch in aller Regel noch nicht erfüllt. Umso wichtiger ist deshalb die Bereitstellung von digitalen Arbeitsumgebungen, mit denen Philologen im ersten Zugriff unstrukturiert oder strukturiert, in jedem Fall aber kollaborativ und damit diskursiv orientiert annotieren, um dann im zweiten Schritt Objekttext wie Annotate mithilfe des Rechners analysieren zu können.

Wichtig ist dabei, den Übergang von der (leichter) automatisierbaren, hermeneutisch jedoch eher niedrigschwelligen formalen Textannotation (z. B. *POS Tagging*) bis hinauf zur nicht mehr algorithmenbasiert reproduzierbaren, dafür jedoch für eine globale Textdeutung entscheidenden semantischen Textannotation als stetig zu denken. Zwischen induktiven und deduktiven, deklarativen und interpretativen Annotationsmodi und automatischen wie intelligenten Verfahren kategorisch zu unterscheiden, hat wenig Sinn, wenn die *DH* mit der traditionellen Texthermeneutik ins Gespräch kommen wollen. Diese grundsätzlichen Kennzeichnungen können jedoch hilfreich sein, um die jeweilige Praxis einer digitalen Textannotation in einem dreidimensionalen Kontinuum zu verorten, dessen Achsen die der Methodik, der Funktion und des Verfahrens sind (s. Abb. 1).

Von einem derart abstrakt formulierten Konzept digitaler Textannotation ausgehend, könnten dann digitale Annotationswerkzeuge und kollaborative Plattformen entwickelt werden, die die konkreten *Use Cases* traditionell arbeitender, hermeneutisch orientierter Forscher aufgreifen, statt wie bisher primär technik- und standardgetrieben voranzuschreiten. Ein strategisch interessanter Ansatz könnte dabei die Formulierung hermeneutischer *Grand Challenges* für die digitale Textannotation und -exploration sein, bei der *DH*-Vertreter und klassisch-hermeneutisch arbeitende Literaturwissenschaftler gemeinsam bewusst Fragestellungen jenseits der linguistisch-textempirischen Phänomene in den Blick nehmen. Von Seiten der Texthermeneutik dürften die formulierten Herausforderungen dabei nur selten darin bestehen, rein technische Routinen und Prozesse der Textannotation als solche zu automatisieren und zu optimieren: Nicht der technologische und informationstheoretische, sondern der epistemologische und konzeptionelle Aspekt dürfte vielmehr bei der Suche nach den wirklichen *Killer Applications*³² im Vordergrund stehen.

³²Patrick Juola, „Killer Applications in Digital Humanities“, in: *Literary and Linguistic Computing* 23/1 (2008), 73–83.

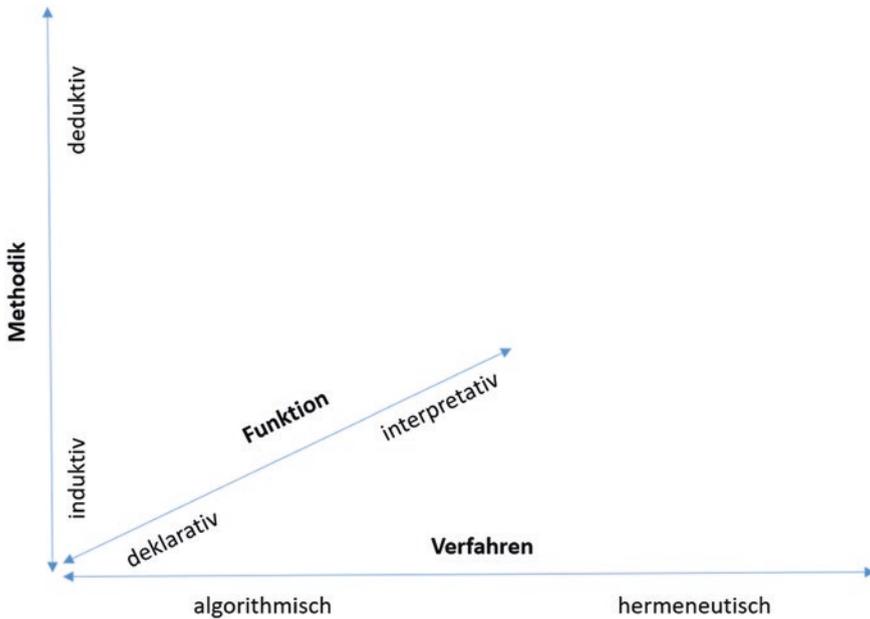


Abb. 1 Dimensionen der digitalen Textannotation

5 Desiderata

Aus den vorangehenden Überlegungen lassen sich drei zentrale Desiderata für die Entwicklung und Dissemination von digitalen Annotationsverfahren und -plattformen ableiten. Anzustreben sind:

- eine umfassende *born-digital*-Modellierung von Textannotation als literaturwissenschaftlichem Verfahren;
- die konzeptionelle und technische Integration der Arbeitsschritte Textauszeichnung – Textannotation – Textanalyse
- eine ‚Naturalisierung‘ digitaler Verfahren in den Philologien.

5.1 *Born-digital*-Modellierung literaturwissenschaftlicher Textannotation

Die digitale Modellierung der traditionellen Praxis non-deterministischer, explorativer Textannotation sollte, wie u. a. Bradley mit Blick auf einen „Richer

Sense of Digital Annotation“³³ hervorgehoben hat, ein erheblich breiteres Spektrum erfassen, als dies mit den gängigen Methoden eines deklarativen, taxonomiegetriebenen Markups möglich ist. Insbesondere die unstrukturierte, intuitive Annotationspraxis muss stärker berücksichtigt werden – was in der traditionellen Praxis bislang auf Papier in Form von Unterstreichungen und Marginalien-Kommentaren zunächst manuell erfasst wurde, um dann allenfalls in einem zweiten Durchgang selektiv in eine digitale Variante kopiert bzw. umgeschrieben zu werden. So sollte also durchgehend eine *born-digital*-Textannotation entstehen können, die dem Nutzer schon in Echtzeit für methodisch anschließende computergestützte Prozesse (Indexierung, *Search and Retrieval*, Schematisierung etc.) zur Verfügung steht.

Dies kann sowohl synchron als auch noch während des Annotationsvorgangs selbst geschehen – etwa durch interaktive Auswertung und Visualisierung der Verteilung bereits annotierter Phänomene im Text- oder Korpus-Ganzen. Unterstützt werden sollte dieser Ansatz *idealiter* jedoch zugleich durch eine asynchrone maschinelle Auswertung des Gesamtkomplexes von Textdaten, Annotationsdaten und im System protokollierten Nutzeraktionen. Ziel einer solchen digitalen Metareflexion unter Verwendung von *Machine-Learning*-Verfahren ist dabei die Modellierung und schrittweise Automatisierung von Annotationsproblemen steigender Komplexität. Dabei geht es nicht allein darum, den Nutzer von repetitiven Aufgaben zu befreien. Durch Einbeziehung der informatischen Perspektive auf Problemstellung und Lösungswege sollten digitale Modellierungsangebote vielmehr auch epistemisch funktional gemacht werden: Denn sie emulieren nicht nur bekannte Verfahren, sondern eröffnen auch neue Perspektiven auf den hermeneutischen Gegenstand ‚Text‘, die bislang aufgrund zu hoher Komplexität für den menschlichen Leser und Kommentator nicht greifbar waren. Wesentlich ist dabei insbesondere die Unterstützung kollaborativer Annotation, die eine methodologische Erweiterung philologischer Annotationspraxis bedeutet.

5.2 Konzeptionelle und technische Integration von Textauszeichnung – Textannotation – Textanalyse

In der traditionellen philologischen Praxis ergänzen sich die Textauszeichnung (Beispiel: eine Analyse des Metrums), die Textannotation (Beispiel: die Explikation einer Metapher), die Textanalyse (die systematische Anwendung der beiden ersten Schritte auf den Gesamttext) und der Textkommentar (Beispiel: die

³³ John Bradley, „Towards a Richer Sense of Digital Annotation. Moving Beyond a ‚Media‘ Orientation of the Annotation of Digital Objects“, in: *DHQ* 6/2 (2012), <http://www.digitalhumanities.org/dhq/vol/6/2/000121/000121.html> (letzter Aufruf 26.4.2018).

kritische Würdigung der Form-Inhalt-Relation zwischen Metrum und Metaphorik vor dem Hintergrund eines Werkganzen). Diese methodische Integration muss auch eine digitale Forschungsumgebung für die Philologien zumindest in Hinblick auf die ersten drei Arbeitsschritte leisten⁶³ – und sie sollte dabei zudem berücksichtigen, dass in der literaturwissenschaftlichen Forschungspraxis gerade diese drei Verfahren eigentlich immer schon diskursiv praktiziert wurden. Denn die jeweils erzielten Ergebnisse fließen seit jeher in den fortlaufenden kritischen Fachdiskurs der Literaturwissenschaften ein. Diskursivität als solche ist also kein Novum – der eigentliche konzeptionelle wie methodische Mehrwert liegt vielmehr in der gesteigerten Komplexität, Synchronizität und Nachnutzbarkeit der ‚Arbeit am Text‘, der aus der digitalen Integration der Verfahren resultiert und im kollaborativen Modus nochmals gesteigert werden kann.

Eine virtuelle Textannotationsumgebung sollte deshalb:

- den Dissens über Bedeutungen als produktive Bedingung und *raison d'être* der Textwissenschaften begreifen;
- den Geltungsbereich der normativen Leitidee des *Inter-Annotator Agreements* bewusst relativieren, indem sie im Umgang mit semantisch anspruchsvolleren Textphänomenen auch das explorative Potenzial eines reflektierten *Disagreements* nutzbar macht;
- statt des linearen Voranschreitens von der taxonomischen Textbeschreibung zur finalen Textdeutung den zyklischen Charakter und die Wechselwirkung von Textannotation und Textanalyse als Prozessschema zugrunde legen.

Für den philologischen Nutzer könnten damit auch in der digitalen Umgebung Textauszeichnung, Textannotation und Textanalyse als methodisches Kontinuum handhabbar bleiben. Die angestrebte *User Experience* emuliert damit die etablierten Praktiken des philologischen Arbeitens, statt Workflows und Segmentierungen von Prozessen zu erzwingen, die sich primär an den technischen Anforderungen orientieren.

5.3 ‚Naturalisierung‘ digitaler Verfahren in den Philologien

Diese Anschlussfähigkeit an etablierte ‚analoge‘ Routinen und Gebrauchsmetaphern ist dabei zugleich ein strategisches Desiderat: Denn mit ihr steht und fällt die Akzeptanz einer Plattform. Ihr wäre mit einem mehrstufigen Usability-Konzept Rechnung getragen, das den konzeptionellen Mehrwert eines *Mixed-Methods*-Ansatzes für die Literaturwissenschaften – also: einer Kombination

⁶³ Das Ausformulieren komplexer Textkommentare bleibt dagegen eine intellektuelle Leistung, die von einer Forschungsumgebung nicht konzeptionell, sondern nur instrumentell unterstützt werden kann – etwa durch Bereitstellung von Recherchetools, Datenbankanbindungen, Kollaborationstools, Workflow-Management etc.

manueller und automatischer Routinen – vermitteln kann. Drei Kriterien sind dabei erfolgskritisch:

- Vom Nutzer dürfen keinerlei Vorkenntnisse im Bereich der digitalen Modellierung und Analyse von Texten und Korpora verlangt und damit vom System auch keine implizite Rekonzeptualisierung des Forschungsgegenstandes ‚Text‘ als Datenstruktur erzwungen werden.
- Im Einstiegs-Modus sollte eine Plattform in Bezug auf Oberflächengestaltung, Terminologie und Handhabung bewusst traditionelle Praktiken, Konzepte, Verfahren und Gebrauchsmetaphern der Philologien emulieren, um dann
- im Experten-Modus schrittweise mittels interaktiver Komponenten zwischen dem traditionellen und dem digital modellierten Workflow zu vermitteln. Wesentlich ist hier insbesondere der Schritt zur kollaborativen Arbeit am Text bzw. Textkorpus, die sowohl asynchron wie synchron, also auch in *Real Time* möglich sein und zudem Gelegenheiten zur Einbindung automatisierter Komponenten bieten muss.

6 CATMA als *Proof of Concept*

Der vorangehend skizzierte Anforderungskatalog mag auf den ersten Blick ambitioniert wirken. Tatsächlich gibt es jedoch bereits jetzt eine Reihe von Tools und Plattformen, die ihm in Konzeption und Funktionalität teilweise entsprechen. Als wohl wichtigstes frühes Beispiel zu nennen ist hier die bereits erwähnte, von Bradley entwickelte experimentelle Annotations- und Forschungsumgebung *Pliny*.⁶⁴ Plinys ausschlaggebende konzeptionelle Restriktion ist allerdings, dass das Tool als Desktopapplikation nur den Arbeitsablauf des Einzelnutzers unterstützt, aber keine Kollaborationsfunktionalität beinhaltet.

Die vorangehend skizzierten konzeptionellen Überlegungen und technischen Erfordernisse sind dagegen umfassender umgesetzt worden in dem seit 2008 an der Universität Hamburg entwickelten Textannotationstool CATMA (Computer Assisted Textual Markup and Analysis).⁶⁵ CATMA war zunächst als Re-Implementierung der bereits in der Mitte der 1980er Jahre von Bradley programmierten Programmsuite TACT (Textual Analysis Computing Tools) geplant,⁶⁶ eines der erfolgreichsten und verbreitetsten frühen Textanalysetools. Im Zuge des Architekturdesigns und nach Rücksprache mit Bradley wurde daraus jedoch ein komplexeres Entwicklungsvorhaben. Neben der Annotationsfunktion wurde das Programm in einer ersten Entwicklungsphase bis 2010 um mehrere

⁶⁴ Bradley (Anm. 26); siehe <http://pliny.cch.kcl.ac.uk> (letzter Aufruf 8.1.2016).

⁶⁵ Siehe <http://www.catma.de> (letzter Aufruf 8.1.2016).

⁶⁶ Ian Lancashire, *Using TACT with electronic texts. A guide to text-analysis computing tools, version 2.1 for MS-DOS and PC DOS*, New York 1996.

Funktionskomponenten erweitert (insbes. Textanalysefunktionen und einen *Query Builder*⁶⁷). Ab 2010 wurde CATMA dann als Webapplikation in JAVA implementiert, um webbasierte Kollaborationen in Echtzeit bei der Text-/Textkorpusannotation und -analyse zu unterstützen. Um die notwendige Flexibilität bei gleichzeitiger Standardkonformität sicherzustellen, basieren CATMA-Annotationen auf dem sog. *Feature-Structure-Modell* von TEI.⁶⁸ CATMA-Annotationen sind damit TEI/XML-kompatibel und können in alle gängigen Formate exportiert werden. Umgekehrt erlaubt CATMA den Import aller gängigen Annotationsformate, insbesondere TEI-XML. Objekttexte und Korpora können in diversen Formaten (darunter txt, rtf, doc, docx, pdf) in das CATMA-System geladen oder über eine Schnittstelle direkt aus Repositorien eingelesen werden. Intern verarbeitet CATMA alle Texte in UTF-8 bzw. UTF-16 und unterstützt somit eine Vielzahl UNICODE-basierter Sprachen inkl. einiger rechts-links geschriebener Sprachen wie z. B. Hebräisch und Arabisch. In CATMA 5.0 wurden außerdem prototypisch erstmals *Machine-Learning*-basierte Annotationsroutinen implementiert, mit denen z. B. das in der Narratologie als *Analepse* definierte Phänomen des non-linearen, rückschauenden Erzählens einer Vorgeschichte vom Zeitpunkt der Erzählgegenwart aus in deutschsprachigen Texten automatisch getaggt werden kann. CATMA wird aktuell international in mehr als 120 Forschungsvorhaben eingesetzt und ist bislang die einzige webbasierte Plattform, die

- gezielt non-deterministische, kollaborative Textannotation unterstützt;
- das komplette Funktionsspektrum ‚Textauszeichnung – Textannotation – Textanalyse‘ abbildet;
- Schnittstellen zur Einbindung externer Routinen und Services für automatisierte Textauszeichnung und Korpora besitzt;
- eine systeminterne *Machine-Learning*-Komponente integriert.

CATMA kann damit als *Proof of Concept* gelten, das die konkrete technische Umsetzbarkeit der methodischen Anforderungen demonstriert, die ein reichhaltigeres, stärker auf die texthermeneutischen Erfordernisse eingehendes *DH*-Annotationskonzept nach sich zieht, für das hier plädiert wurde.⁶⁹ CATMA ist zwar bislang die einzige Plattform, die dem entsprechenden technischen Anforderungskatalog gerecht wird.⁷⁰ Aber das ist durchaus keine Konsequenz

⁶⁷ D. i. eine natürlichsprachliche Oberfläche zur Generierung komplexer Anfragen sowie einer Datenvisualisierung.

⁶⁸ Zum *Feature-Structure*-Konzept siehe <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/FS.html> (letzter Aufruf 8.1.2016).

⁶⁹ CATMA wird in dem aktuellen DFG-LIS-Projekt *forTEXT*, <http://www.fortext.net> (letzter Aufruf 8.1.2018) schrittweise um u. a. folgende Komponenten erweitert werden: Skalierbarkeit & Performanz; Rollen & Rechte-Schemata sowie Usability-Anforderungen; Aufbau eines ML-basierten Recommender-Funktionskomplexes; Integration von Visualisierungsfunktionen; UIMA-basierte Einbindung weiterer externer Services; Anbindung an Textrepositorien wie z. B. *TextGrid* oder *CLARIN*-Repositorien.

⁷⁰ Siehe hierzu die Bestandsaufnahme im Anhang, die im Rahmen der Antragstellung zum Projekt *forTEXT* von Marco Petris erarbeitet wurde.

technischer Restriktionen, die auf das Feld der *DH* einwirken. Es ist vielmehr eine Folge der konzeptionellen Selbstbeschränkung der *DH* auf eine Praxis der digitalen Anreicherung von Texten mit ‚Lesartenhinweisen‘, die Markup bleiben. Das aber ist nur die Vorstufe zu dem, was in der Textthermeneutik eigentlich gefordert ist: digitale Annotation.

Anhang

Im Folgenden werden die Ergebnisse der Bestandsaufnahme und Evaluation bestehender Plattformen und Werkzeuge für die Textannotation und -analyse dargestellt, die 2016 für den Antrag zum DFG-geförderten LIS-Projekt *forTEXT* erarbeitet wurden (Autor: Marco Petris). Dabei wurden ausschließlich Plattformen berücksichtigt, die aktiv entwickelt und betreut werden.⁷¹ Die Plattformen wurden bezüglich ihrer Funktionalität, Zielgruppe und Nachhaltigkeit betrachtet. Hinsichtlich der Zielgruppenspezifität wurde insbesondere die erforderliche technische Versiertheit potenzieller Nutzer bewertet. Abhängig von technischem Vorverständnis und Interesse lassen sich dabei drei Nutzertypen unterscheiden:

1. Nutzer ohne Bereitschaft, sich mit technischem Hintergrundwissen auseinanderzusetzen;
2. Nutzer mit der Bereitschaft, sich mit technischem Hintergrundwissen auseinanderzusetzen;
3. Nutzer mit der Bereitschaft, auf Programmcode-Ebene zu arbeiten.

Kriterien für Nachhaltigkeit der untersuchten Plattformen sind i) das Setzen auf offene Standards und ii) die Veröffentlichung des Programmcodes unter einer die Nachnutzbarkeit ermöglichenden Lizenz. Die Ergebnisse der Bestandsaufnahme werden hier tabellarisch dargestellt (s. Tab. A.1). Die Gesamtauswertung ergab, dass 2016 von den bislang entwickelten Werkzeugen keines die in *forTEXT* anvisierten Funktionalitäten im Sinne eines *Hermeneutic-Markup*-Ansatzes abdeckte und dabei

- mit wenig technischem Hintergrundwissen bedienbar ist,
- Funktionen zur Annotation und zur Auswertung integriert,
- kollaboratives Arbeiten ermöglicht,
- die Möglichkeit zur Modellierung komplexerer Taxonomien bereitstellt und
- den Quellcode zur Verfügung stellt.

⁷¹ Projekte wie *Project Bamboo* oder die *Open Corpus Workbench* spielen deshalb als mögliche Komponenten von *forTEXT* keine Rolle, siehe auch: Mark A. Finlayson/Patrick H. Winston, *The Rapid Story Annotation Workbench* (2007), <http://publications.csail.mit.edu/abstracts/abstracts07/markaf/markaf.html> (letzter Aufruf 8.6.2013).

Tab. A.1 Bestandsaufnahme und Evaluation bestehender Plattformen und Werkzeuge für die Textannotation und -analyse

Plattform	Hauptfunktionen	Source Code veröffentlicht? (Lizenz)	Nutzertyp
Apache UIMA ³⁴	<ul style="list-style-type: none"> • Java/C++-basierte Komponenten können über die Common Analysis Structure (CAS) zu Workflows zusammengefasst und über eine Workflow Engine ausgeführt werden • vertikale und horizontale Skalierung • zahlreiche Komponenten für die Arbeit mit Texten • erlaubt die Definition von Typsystemen für die Erzeugung/Verarbeitung von Analysedaten 	ja (ASL) ³⁵	3
DKPro Core ³⁶ & DKPro Lab ³⁷	<p>DKPro Core:</p> <ul style="list-style-type: none"> • Sammlung von Komponenten für NLP auf Basis von Apache UIMA <p>DKPro Lab:</p> <ul style="list-style-type: none"> • Framework zum Erstellen von deklarativ in Java notierten Workflows • Workflows mit unterschiedlichen Parameterbelegungen ausführbar (Fokus auf Effizienz: Ein einzelner Schritt wird nur einmal ausgeführt, sofern sich die Parameter für diesen Schritt nicht ändern.) • speichert Ein- und Ausgabewerte des Workflows, deshalb stehen die Informationen (<i>Data Provenance</i>) auch nach Ausführung noch in vollem Umfang zur Verfügung 	ja (ASL, Teile von DKPro Core in GPL ³⁸)	3
GATE ³⁹	<ul style="list-style-type: none"> • Java-basierte Komponenten können zu Workflows zusammengefasst werden • Komponenten erzeugen bzw. verarbeiten Annotationen mit frei definierbaren Typsystemen 	Ja	3 (GATE Developer: 2)

(Fortsetzung)

³⁴UIMA (Unstructured Information Management Architecture), <https://uima.apache.org/> (letzter Aufruf 8.1.2016). In vorangegangenen Projekten konnten bereits Komponenten zur Integration von CATMA und UIMA erarbeitet werden. Hierzu zählen die Anbindung an die CATMA API zum Import und Export von Texten und Annotationen, der Collection Reader zur Verarbeitung von CATMA Source Documents und die Abbildung von UIMA-Typsystemen auf CATMA-Typsysteme.

³⁵*Apache Software License* (vgl. <http://www.apache.org/licenses/> [letzter Aufruf 8.1.2016]).

³⁶Vgl. Richard E. de Castilho/Iryna Gurevych, „A Broad-Coverage Collection of Portable NLP Components for Building Shareable Analysis Pipelines“, in: *Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT (2014)*, 1–11.

³⁷Vgl. Richard E. de Castilho/Iryna Gurevych, „A Lightweight Framework for Reproducible Parameter Sweeping in Information Retrieval“, in: *DESIRE (2011)*, 7–10.

³⁸*GNU Public License* (vgl. <http://www.gnu.org/licenses/gpl-3.0.de.html> [letzter Aufruf 8.1.2016]).

³⁹*General Architecture for Text Engineering* (vgl. Hamish Cunningham/Kalina Bontcheva/Valentin Tablan u. a. [Hg.], *Text Processing with GATE. [Version 6]*, Sheffield 2011).

Tab. A.1 (Fortsetzung)

Plattform	Hauptfunktionen	Source Code veröffentlicht? (Lizenz)	Nutzertyp
	<ul style="list-style-type: none"> • umfangreiche Sammlung an Komponenten • GATE-Komponenten können in UIMA-Workflows und UIMA-Komponenten in GATE-Workflows integriert werden • GATE Developer: Desktopwerkzeug zum Erzeugen und Ausführen von Workflows und zum manuellen Annotieren von Texten • GATE Teamware: Webapplikation für Management von Projekten und Teams • GATE Server: ermöglicht kollaboratives Arbeiten mit dem GATE Developer • Lucene⁴²-basierte Suche und JAPE⁴³ • Möglichkeit, über bestehende Annotationen durch reguläre Ausdrücke neue Annotationen zu vergeben 	(LGPL ⁴⁰ , GATE Teamware: AGPL ⁴¹)	
KNIME ⁴⁴ Taverna ⁴⁵ Kepler ⁴⁷	<ul style="list-style-type: none"> • Plattform zum Erstellen und Ausführen von naturwissenschaftlichen Workflows • <i>Rich-Client</i>-Anwendungen mit graphischem Designer zum Erstellen von Workflows aus Komponenten • Komponenten für naturwissenschaftliche Workflows (kleinteiliger, deshalb flexibler nutzbar) 	ja (GPL, ASL, BSD ⁴⁶)	2

(Fortsetzung)

⁴⁰ *Lesser GNU Public License* (vgl. <http://www.gnu.de/documents/lgpl-3.0.de.html> [letzter Aufruf 8.1.2016]).

⁴¹ *Affero GNU Public License* (vgl. <http://www.gnu.org/licenses/agpl-3.0.de.html> [letzter Aufruf 8.1.2016]).

⁴² Vgl. <http://lucene.apache.org/> (letzter Aufruf 8.1.2016).

⁴³ Vgl. Hamish Cunningham/Diana Maynard/Valentin Tablan, *JAPE. A Java Annotation Patterns Engine*, o. O. 2000, https://www.researchgate.net/publication/2495768_JAPE_a_Java_Annotation_Patterns_Engine/citation/download (letzter Aufruf 8.1.2016).

⁴⁴ Vgl. Michael R. Berthold/Nicolas Cebron/Fabian Dill u. a., „KNIME. The Konstanz Information Miner“, in: Christine Preisach/Hans Burkhardt/Lars Schmidt-Thieme u. a. (Hg.), *Data Analysis, Machine Learning and Applications*, Berlin/Heidelberg 2008, 319–26, http://link.springer.com/10.1007/978-3-540-78246-9_38, (letzter Aufruf 3.5.2018).

⁴⁵ Vgl. Katherine Wolstencroft/Robert Haines/Donal Fellows u. a., „The Taverna Workflow Suite. Designing and Executing Workflows of Web Services on the Desktop, Web or in the Cloud“, in: *Nucleic Acids Research* 41 (2013), W557–W561.

⁴⁶ Vgl. <https://opensource.org/licenses/BSD-2-Clause> (letzter Aufruf 8.1.2016).

⁴⁷ Vgl. Ilkay Altintas/Chad Berkley/Efrat Jaeger u. a., „Kepler. An extensible system for design and execution of scientific workflows“, in: *Proceedings of the 16th International Conference on Scientific and Statistic Database Management* (2004), 21–23.

Tab. A.1 (Fortsetzung)

Plattform	Hauptfunktionen	Source Code veröffentlicht? (Lizenz)	Nutzertyp
	<ul style="list-style-type: none"> • externe Software einbindbar (z. B. R⁴⁸ oder Weka⁴⁹) • Workflows können exportiert und über Seiten wie myExperiment⁵⁰ geteilt werden 		
MAXQDA ⁵¹ ATLAS.ti ⁵²	<ul style="list-style-type: none"> • <i>Rich-Client</i>-Anwendungen zur qualitativen Datenanalyse • Schwerpunkt auf sozialwissenschaftlichen Problemstellungen • Modellierung von Taxonomien wegen eindimensionaler Annotationstypen (<i>Codes</i>) 	nein (kommerzielle Lizenzen)	1
TextGrid ⁵³	<ul style="list-style-type: none"> • dokumentierte Schnittstellen zur Erweiterung bzw. zur Integration mit anderen Systemen • unterstützt kollaboratives Arbeiten in Teams mithilfe einer projektbezogenen Benutzerverwaltung auf Basis von openRBAC⁵⁴ <p>TextGridRepository:</p> <ul style="list-style-type: none"> • verteilte Speicherinfrastruktur für Datenhaltung • TextGrid Middleware Utilities (kapselt Zugriff auf den Speicher für die Services-Schicht) <p>TextGridLaboratory:</p> <ul style="list-style-type: none"> • Services bzw. Komponenten für die Analyse und Verarbeitung von Daten⁵⁵ 	ja (LPGL)	1
	<ul style="list-style-type: none"> • graphische Schnittstelle zum Endnutzer basierend auf Eclipse RCP⁵⁶, erweiterbar durch Plugins 		

(Fortsetzung)

⁴⁸Vgl. <https://www.r-project.org/> (letzter Aufruf 8.1.2016).

⁴⁹Vgl. <http://www.cs.waikato.ac.nz/ml/weka/> (letzter Aufruf 8.1.2016).

⁵⁰Vgl. www.myexperiment.org/ (letzter Aufruf 8.1.2016).

⁵¹Vgl. <http://www.maxqda.de/> (letzter Aufruf 8.1.2016).

⁵²Vgl. <http://atlasti.com> (letzter Aufruf 8.1.2016).

⁵³Vgl. Heike Neuroth/Andrea Rapp/Sibylle Söring u. a. (Hg.), *TextGrid. Von der Community – für die Community*, Göttingen 2015.

⁵⁴Vgl. http://www.openrbac.de/en_startup.xml (letzter Aufruf 8.1.2016).

⁵⁵Zurzeit sind für *TextGrid* vergleichsweise wenig Komponenten direkt verfügbar. Anwendungsfälle aus der Editionsphilologie werden durch die bestehenden Komponenten jedoch bereits unterstützt. Die Unterstützung zum Modellieren von Workflows ist noch in einem experimentellen Stadium und in der aktuellen Version (Nightly Build 3.1.0.201507151344) nicht enthalten bzw. wird nicht über den *TextGrid Marketplace* angeboten.

⁵⁶Vgl. <http://www.eclipse.org> (letzter Aufruf 8.1.2016).

Tab. A.1 (Fortsetzung)

Plattform	Hauptfunktionen	Source Code veröffentlicht? (Lizenz)	Nutzertyp
WebAnno ⁵⁷ und ANNIS ⁵⁸	<p>WebAnno:</p> <ul style="list-style-type: none"> • kollaboratives, webbasiertes Annotationswerkzeug • eigene Annotationstypen definierbar (Modellierungsmöglichkeit ist aufgrund der eindimensionalen Typen begrenzt; kleinste annotierbare Zeichenkette ist ein Wort) <p>ANNIS:</p> <ul style="list-style-type: none"> • Analysewerkzeug zur Visualisierung und Auswertung in WebAnno erstellter Annotationen (allerdings kein integrierter Ansatz) 	ja (ASL)	1
WebLicht ⁵⁹	<ul style="list-style-type: none"> • Komponenten auf Basis von REST Web Services können zu Workflows zusammengefasst und ausgeführt werden • eigenes Datenformat: TCF⁶⁰ • umfangreiche Bibliothek mit Komponenten • über Web UI benutzbar und als Web-Service in Plattformen wie GATE und UIMA einbindbar⁶¹ • mit Suchkomponente können Anfragen in CQL⁶² gestellt werden 	nein (steht Forschern und Studierenden zur Verfügung)	2

⁵⁷ Vgl. Seid Muhie Yimam/Iryna Gurevych/Richard E. de Castilho u. a., „A Flexible, Web-based and Visually Supported System for Distributed Annotations“, in: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (2013), 1–6.

⁵⁸ Vgl. Michael Götz/Stefanie Dipper, „ANNIS, Complex Multilevel Annotations in a Linguistic Database (System Demonstration)“, in: *Proceedings of the EACL Workshop on Multidimensional Markup in Natural Language Processing NLPXML* (2006), 61–64.

⁵⁹ Web-Based Linguistic Chaining Tool, vgl. Erhard Hinrichs/Marie Hinrichs/Thomas Zastrow, „WebLicht. Web-based LRT Services for German“, in: *Proceedings of the ACL 2010 System Demonstrations* (2010), 25–29.

⁶⁰ Ulrich Heid/Helmut Schmid/Kerstin Eckart u. a., „A Corpus Representation Format for Linguistic Web Services. The D-SPIN Text Corpus Format and its Relationship with ISO Standards“, in: *Proceedings of the International Conference on Language Resources and Evaluation, LREC* (2010), <http://www.lrec-conf.org/proceedings/lrec2010/summaries/503.html> [letzter Aufruf 3.5.2018].

⁶¹ Das Einbinden von eigenen Komponenten in *WebLicht* ist jedoch vergleichsweise aufwändig. Diese müssen via CMDI (<http://www.clarin.eu/cmdi>) beschrieben und über ein CLARIN-D-Repository, welches Web Services unterstützt (<http://www.clarin-d.de/de/aufbereiten/clarin-zentrum-finden> [letzter Aufruf der Links 8.1.2016]), veröffentlicht werden.

⁶² Oliver Christ, „A Modular and Flexible Architecture for an Integrated Corpus Query System“, in: *Proceedings of COMPLEX, Conference on Computational Lexicography and Text Research* (1994), 23–32.

Literatur

- Altintas, Ilkay/Berkley, Chad/Jaeger, Efrat u.a., „Kepler. An extensible system for design and execution of scientific workflows“, in: *Proceedings of the 16th International Conference on Scientific and Statistic Database Management* (2004), 21–23.
- Bauer, Matthias/Zirker, Angelika, „Whipping Boys Explained. Literary Annotation and Digital Humanities“, in: Kenneth M. Price/Ray Siemens (Hg.), *Literary Studies in the Digital Age. An Evolving Anthology*, New York 2013, <https://dlsanthology.mla.hcommons.org/> (letzter Aufruf 3.5.2018).
- Berthold, Michael R./Cebron, Nicolas/Dill, Fabian u.a., „KNIME. The Konstanz Information Miner“, in: Christine Preisach/Hans Burkhardt/Lars Schmidt-Thieme u.a. (Hg.), *Data Analysis, Machine Learning and Applications*, Berlin/Heidelberg 2008, 319–26, http://link.springer.com/10.1007/978-3-540-78246-9_38, (letzter Aufruf 3.5.2018).
- Boot, Peter, *Mesotext. Digitised Emblems, Modelled Annotations and Humanities Scholarship*, Utrecht 2009.
- Bradley, John, „Thinking about interpretation. Pliny and scholarship in the humanities“, in: *Literary and Linguistic Computing* 23 (2008), 263–279.
- Bradley, John, „Towards a Richer Sense of Digital Annotation. Moving Beyond a ‚Media‘ Orientation of the Annotation of Digital Objects“, in: *DHQ* 6/2 (2012), <http://www.digitalhumanities.org/dhq/vol/6/2/000121/000121.html> (letzter Aufruf 26.4.2018).
- Buzzetti, Dino, „Digital Representation and the Text Model“, in: *New Literary History* 33 (2002), 61–88.
- Buzzetti, Dino, „Text Representation and Textual Models“, in: *1999 ACH-ALLC Conference Program*, <http://www2.iath.virginia.edu/ach-allc.99/proceedings/buzzetti.html> (letzter Aufruf 7.5.2018)
- Castilho, Richard E. de/Gurevych, Iryna, „A Broad-Coverage Collection of Portable NLP Components for Building Shareable Analysis Pipelines“, in: *Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT* (2014), 1–11.
- Castilho, Richard E. de/Gurevych, Iryna, „A Lightweight Framework for Reproducible Parameter Sweeping in Information Retrieval“, in: *DESIRE* (2011), 7–10.
- Caton, Paul, „Markup’s Current Imbalance“, in: *Markup Languages* 3/1 (2000), 1–13.
- Christ, Oliver, „A Modular and Flexible Architecture for an Integrated Corpus Query System“, in: *Proceedings of COMPLEX, Conference on Computational Lexicography and Text Research* (1994), 23–32.
- Coombs, James H./Renear, Allen H./DeRose, Steven J., „Markup systems and the future of scholarly text processing“, in: *Communications of the ACM* 30 (1987), 933–947.
- Cunningham, Hamish/Bontcheva, Kalina/Tablan, Valentin (Hg.), *Text Processing with GATE. (Version 6)*, Sheffield 2011.
- Cunningham, Hamish/Maynard, Diana/Tablan, Valentin, *JAPE. A Java Annotation Patterns Engine*, o.O. 2000, https://www.researchgate.net/publication/2495768_JAPE_a_Java_Annotation_Patterns_Engine/citation/download (letzter Aufruf 7.5.2018).
- Dombrowski, Quinn, „What Ever Happened to Project Bamboo?“, in: *Literary and Linguistic Computing* 29/3 (2014), 326–339.
- Finlayson, Mark A./Winston, Patrick H., *The Rapid Story Annotation Workbench* (2007), <http://publications.csail.mit.edu/abstracts/abstracts07/markaf/markaf.html> (letzter Aufruf 8.6.2013).
- Flanders, Julia/Jannidis, Fotis, „Data Modeling“, in: Susan Schreibman/Ray Siemens/John Unsworth (Hg.), *A New Companion to Digital Humanities*, Chichester, West Sussex, UK, 2016, 229–237.
- Götze, Michael/Dipper, Stefanie, „ANNIS, Complex Multilevel Annotations in a Linguistic Database (System Demonstration)“, in: *Proceedings of the EACL Workshop on Multi-dimensional Markup in Natural Language Processing NLPXML* (2006), 61–64.

- Heid, Ulrich/Schmid, Helmut/Eckart, Kerstin u.a., „A Corpus Representation Format for Linguistic Web Services. The D-SPIN Text Corpus Format and its Relationship with ISO Standards“, in: *Proceedings of the International Conference on Language Resources and Evaluation, LREC* (2010), <http://www.lrec-conf.org/proceedings/lrec2010/summaries/503.html> (letzter Aufruf 3.5.2018).
- Hinrichs, Erhard/Hinrichs, Marie/Zastrow, Thomas, „WebLicht. Web-based LRT Services for German“, in: *Proceedings of the ACL 2010 System Demonstrations* (2010), 25–29.
- Hockey, Susan/McGann, Jerome J./Renear, Allen, „What is text? A debate on the philosophical and epistemological nature of text in the light of humanities computing research“, in: *1999 ACH-ALLC Conference Program*, <http://www2.iath.virginia.edu/ach-allc.99/proceedings/hockey-renear2.html> (letzter Aufruf 7.5.2018).
- Jacke, Janina, „Is There a Context-Free Way of Understanding Texts? The Case of Structuralist Narratology“, in: *Journal of Literary Theory* 8/1 (2014), 118–139.
- Juola, Patrick, „Killer Applications in Digital Humanities“, in: *Literary and Linguistic Computing* 23/1 (2008), 73–83.
- Lancashire, Ian, *Using TACT with electronic texts. A guide to text-analysis computing tools, version 2.1 for MS-DOS and PC DOS*, New York 1996.
- Marcoux, Yves/Sperberg-McQueen, Michael/Huitfeldt, Claus, „Modeling overlapping structures“, in: *Balisage. The Markup Conference* (2013), <https://www.balisage.net/Proceedings/vol10/html/Marcoux01/BalisageVol10-Marcoux01.html> (letzter Aufruf 26.4.2018).
- Mayer-Schönberger, Viktor/Cukier, Kenneth, *Big data. A revolution that will transform how we live, work, and think*, Boston 2013.
- McGann, Jerome, „Marking texts in many dimensions.“, in: *A new republic of letters: Memory and scholarship in the age of digital reproduction*, Cambridge, MA, 2014, 90–112.
- Meister, Jan C. u. a., *DARIAH-DE Folgeantrag, Arbeitspaket 5.4*, internes Dokument 2013.
- Neuroth, Heike/Rapp, Andrea/Söring, Sibylle u. a. (Hg.), *TextGrid. Von der Community – für die Community*, Göttingen 2015.
- Pierazzo, Elena, „Textual Scholarship and Text Encoding“, in: Susan Schreibman/Ray Siemens/John Unsworth (Hg.), *A New Companion to Digital Humanities*, Chichester, West Sussex u.a. 2016, 307–321.
- Piez, Wendell, *Towards Hermeneutic Markup. An architectural outline* (2010), <http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/html/ab-743.html> (letzter Aufruf 9.11.2017).
- Rapp, Andrea, „Manuelle und automatische Annotation“, in: Fotis Jannidis/Hubertus Kohle/Malte Rehbein (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart 2017, 253–267.
- Renear, Allen, „Text Encoding“, in: Susan Schreibman/Ray Siemens/John Unsworth (Hg.), *Companion to Digital Humanities*, Oxford 2004.
- Renear, Allen, „The descriptive/procedural distinction is flawed“, in: *Markup Languages* 2/4 (2000), 411–420.
- Schloen, David/Schloen, Sandra, „Beyond Gutenberg. Transcending the Document Paradigm in Digital Humanities“, in: *DHQ* 8/4 (2014), <http://www.digitalhumanities.org/dhq/vol/8/4/000196/000196.html> (letzter Aufruf 3.5.2018).
- Staten, Henry, „Wittgenstein and the Intricate Evasions of ‚Is‘“, in: *New Literary History* 19/2 (1988), 281–300.
- Unsworth, John, *Scholarly Primitives. What methods do humanities researchers have in common, and how might our tools reflect this?*, London 2000.
- Nyhan, Julianne, „Text encoding and scholarly digital editions“, in: Dies./Claire Warwick/Melissa Terras, *Digital Humanities in Practice. Facet Publishing*, London 2012, 117–138.
- Witt, Andreas, „Multiple hierarchies. New aspects of an old solution“, in: *Proceedings of Extreme Markup Languages* (2004), <http://conferences.idealliance.org/extreme/html/2004/Witt01/EML2004Witt01.html> (letzter Aufruf 8.1.2017).
- Wittgenstein, Ludwig, *Philosophische Untersuchungen*, Frankfurt a. M. 1975 [1953].

- Wolstencroft, Katherine/Haines, Robert/Fellows, Donal u.a., „The Taverna Workflow Suite. Designing and Executing Workflows of Web Services on the Desktop, Web or in the Cloud“, in: *Nucleic Acids Research* 41 (2013), W557–W561.
- Yimam, Seid Muhie/Gurevych, Iryna/de Castilho, Richard E. u. a., „A Flexible, Web-based and Visually Supported System for Distributed Annotations“, in: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (2013), 1–6.
- Zundert, Joris J. van, „Screwmenetics and Hermenumerals. The Computationality of Hermeneutics“, in: Susan Schreibman/Ray Siemens/John Unsworth (Hg.), *A New Companion to Digital Humanities*, Chichester, West Sussex, UK, u. a. 2016.

Online-Ressourcen⁷²

- 3DH, Jan C. Meister, <http://jcmeister.de/projects/3dh/>.
- Apache Licenses, <http://www.apache.org/licenses/>
- Apache Lucene, <http://lucene.apache.org/>.
- Apache UIMA, <https://uima.apache.org/>.
- ATLAS.ti, <http://atlasti.com>.
- CATMA 5.0, <http://www.digitalhumanities.it/catma>.
- CATMA, <http://www.catma.de>.
- CLARIN-D, Zentren, <http://www.clarin-d.de/de/aufbereiten/clarin-zentrum-finden>.
- Database System Research, Michael Gertz, <http://dbs.ifi.uni-heidelberg.de/index.php?id=129>.
- DHd – Digital Humanities im deutschsprachigen Raum, www.dig-hum.de.
- Eclipse, <http://www.eclipse.org>.
- forTEXT, <http://www.fortext.net>.
- GitHub, mpetris [d.i. Marco Petris], CATMA, <https://github.com/mpetris/catma/issues?q=is%3Aopen+is%3Aissue+label%3Aenhancement>.
- GNU, Affero General Public License, <http://www.gnu.org/licenses/agpl-3.0.de.html>.
- GNU, General Public License 3, EN, <http://www.gnu.org/licenses/gpl-3.0.de.html>.
- GNU, Lesser General Public License 3, DE, <http://www.gnu.de/documents/lgpl-3.0.de.html>.
- heureCLÉA, <http://heureclea.de>.
- MAXQDA, <http://www.maxqda.de/>.
- MyExperiment, <http://www.myexperiment.org/>.
- NINES, Nineteenth-Century Scholarship Online, <http://www.nines.org>.
- Open RBAC, http://www.openrbac.de/en_startup.xml.
- Open Source Initiative, The 2-Clause BSD License, <https://opensource.org/licenses/BSD-2-Clause>.
- Pliny, <http://pliny.cch.kcl.ac.uk>.
- Project Bamboo, <http://www.projectbamboo.org/>.
- Text Encoding Initiative (TEI), <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/FS.html>.
- The R Project for Statistical Computing, <https://www.r-project.org/>.
- WebLicht, FAQ, http://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/FAQ#Can_I_use_WebLicht_even_if_my_institute_is_not_on_the_provider_li st.3F.
- Weka 3, Machine Learning Software in Java, <http://www.cs.waikato.ac.nz/ml/weka/>.

⁷² Sämtliche zitierten Online-Ressourcen wurden letztmalig am 8.1.2016 eingesehen.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Digitale Hermeneutik. Computergestütztes *Close Reading* als literaturwissenschaftliches Forschungsparadigma?

Evelyn Gius

1 Digitale und nicht-digitale Literaturwissenschaft

Die Frage nach dem Status der *Digital Humanities* hat mittlerweile durch ihre Institutionalisierung in Form von Professuren, Forschungsfeldern, Konferenzen und Förderlinien eine pragmatische Antwort gefunden oder zumindest deutlich an Dringlichkeit verloren. Damit geht eine Phase teilweise ermüdender und wenig ergiebiger Diskussionen zu Ende. Die Diskussionen um die Inhalte der *Digital Humanities* sind im Umfang deutlich weniger, vielleicht auch, weil man sie – lapidar, aber doch treffend – abkürzen kann: Im Grunde handelt es sich immer dann um *Digital Humanities*, wenn der Gegenstand oder die Methodik einer geisteswissenschaftlichen Forschung digital ist. Allerdings scheint es im Kontext der disziplinären Entwicklungen einen wesentlichen Unterschied zu machen, welchen Teil nun das Digitale ausmacht. Während nämlich digitale Gegenstände relativ problemlos in die traditionellen Geisteswissenschaften integriert worden sind, wurden und werden digitale Zugänge bzw. Methoden weniger offen aufgenommen. Die damit verbundene Frage nach dem Verhältnis der nicht-digitalen Geisteswissenschaften zu den *Digital Humanities* ist durchaus legitim. Ihre Diskussion sollte aber nicht als Grundsatzdebatte geführt werden, an deren Ende es nur einen ‚richtigen‘ Zugang geben kann, sondern vielmehr als gewinnbringende Perspektive auf die beteiligten Seiten genutzt werden. Durch die Betrachtung beider Felder und ihrer Ähnlichkeiten und Unterschiede werden nämlich Eigenarten der Felder sichtbar, die ohne die Kontrastierung kaum wahrgenommen werden, jedoch zur Reflexion und damit zur Weiterentwicklung beider Felder beitragen können.

E. Gius (✉)
TU Darmstadt, Darmstadt, Deutschland
E-Mail: evelyn.gius@tu-darmstadt.de

Dieser Beitrag will sich in einem Bereich der Literaturwissenschaft an einer solchen kontrastierenden Betrachtung versuchen. Im Zentrum soll ein Aspekt der literaturwissenschaftlichen Arbeit stehen, der ein, wenn nicht *das* literaturwissenschaftliche Verfahren par excellence ist: die hermeneutische Analyse literarischer Texte. Diese soll beispielhaft für das literaturwissenschaftliche Forschungsparadigma betrachtet werden. Dabei sollen als ‚hermeneutisch‘ jene Verfahren bezeichnet werden, die auf das Lesen, Verstehen und Interpretieren von Texten ausgerichtet sind.¹ Dieses paradigmatische literaturwissenschaftliche Verfahren bildet den Hintergrund, vor dem die Erfahrungen aus zwei *Digital-Humanities*-Projekten beleuchtet werden sollen, die mit textnahen, manuellen Annotationen gearbeitet haben. Die Kontrastierung besteht entsprechend nicht in der Gegenüberstellung der nicht-digitalen und der digitalen Zugangsweise, sondern darin, die digitale Zugangsweise auf die Verfahren der nicht-digitalen Literaturwissenschaft zu beziehen. Aufgrund der methodischen Vielfalt in der Literaturwissenschaft ist eine direkte Gegenüberstellung beider Zugänge schwierig. Die Betrachtung kann deshalb hauptsächlich Aussagen darüber hervorbringen, ob und wo die digitale Zugangsweise die etablierte Literaturwissenschaft verändert, ergänzt oder erweitert.²

Diese Rückbindung an die nicht-digitale Literaturwissenschaft erscheint aber erfolgsversprechend, da die manuelle Annotation von Texten ein Verfahren ist, das – im Sinne des Hervorhebens und ggf. Kommentierens bzw. Interpretierens von Textstellen – auch unabhängig von *Digital-Humanities*-Kontexten als *Close-*

¹Klaus Weimar, „Hermeneutik“, in: Harald Fricke/Klaus Grubmüller/Jan-Dirk Müller u. a. (Hg.), *Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*, Berlin 2000, 25–29, hier: 25.

²Eine Möglichkeit für eine direkte Gegenüberstellung wäre die Durchführung eines Experiments, bei dem mit denselben Methoden digital und nicht-digital – parallel und unabhängig voneinander – zur selben Forschungsfrage und demselben Gegenstand gearbeitet wird und die stattfindenden Prozesse dokumentiert werden. Dies wäre im Prinzip eine wiederholende Forschung im Sinne einer Replikation – vgl. Christof Schöch, „Wiederholende Forschung in den digitalen Geisteswissenschaften“, in: *DHd 2017. Digitale Nachhaltigkeit. Konferenzabstracts* (2017), 207–212, hier: 208, http://www.dhd2017.ch/wp-content/uploads/2017/02/Abstractband_ergaenzt.pdf (letzter Aufruf 15.9.2017) –, bei der allerdings die Wiederholung des Experiments praktisch zeitgleich zur Durchführung stattfinden würde. Wiederholende Forschungen sind in den *Digital Humanities* durchaus üblich. Geht man davon aus, dass eine Wiederholung nicht-digitaler Forschung mit einem algorithmischen Zugang automatisch eine veränderte Methode bedeutet, handelt es sich allerdings nicht um eine Replikation, sondern um eine Reanalyse (ebd.). Selbst wenn man den algorithmischen Zugang nicht als neue Methode bewertet: Eine Replikation, die neben der gleichen Fragestellung und denselben Daten auch dieselbe Methodik nutzt, steht meistens vor dem Problem nicht dokumentierter Elemente des Forschungsprozesses – wie fehlender Definitionen von Konzepten für die Anwendung, impliziter Annahmen und ungenauer Angaben zur konkreten Anwendung. Deshalb muss mit Annahmen über diese fehlenden Elemente gearbeitet werden, wodurch eine tatsächliche Wiederholung und damit ein methodologischer Vergleich der Zugänge unmöglich wird (vgl. dazu auch die Ausführungen in dem erwähnten Aufsatz von Schöch).

Reading-Technik betrachtet wird.³ Das betrachtete textanalytische Verfahren ist also ursprünglich ein nicht-digitales, weshalb der Hiatt zwischen traditionellem Methodenparadigma und den angewendeten Methoden wesentlich kleiner sein sollte als im Falle von auf meist wesentlich größere Textmengen *Distant-Reading-Methoden* wie stilometrischen Analysen oder *Topic Modeling*. Durch diese größere Nähe zur etablierten literaturwissenschaftlichen Praxis der Textanalyse kann der Zugang auch jene Literaturwissenschaftlerinnen und Literaturwissenschaftler ansprechen, die eben doch eine „fitness of method“, eine methodische Passung zum Gegenstand, erwarten und an der „determination of interpretive boundaries“ interessiert sind – Aspekte, die Stephen Ramsay im Kontext des *Algorithmic Criticism* im Rahmen digitaler Zugänge für wenig relevant hält, die aber fest im Selbstverständnis der traditionellen Literaturwissenschaft verankert sind.⁴

2 *Close Reading in den Digital Humanities*

Während in der literaturwissenschaftlichen Textanalyse *Close-Reading-Verfahren* durchaus etabliert sind, stellen sie im Bereich der *Digital Humanities* eher eine Ausnahme dar. Automatisierte Verfahren werden nicht für das *Close Reading* genutzt, weil sie dafür aus philologischer Sicht (noch) nicht geeignet sind, aber auch Verfahren wie das der manuellen und damit analyseintensiven Annotation werden bislang – außerhalb des Feldes der Editionsphilologie oder zur Anreicherung von Korpora für die maschinelle Verarbeitung – selten angewendet. Insbesondere die interpretationsbasierte Annotationstätigkeit kommt im *Digital-Humanities-Kontext* kaum vor. Entsprechend wird in der *Taxonomy of Digital Research Activities in the*

³ *Close Reading* steht dabei weniger für das vom *New Criticism* propagierte Verfahren der werkimmanenten Interpretation in der von I. A. Richards (*Practical Criticism* und *The Meaning of Meaning*) und William Empson (*Seven Types of Ambiguity*) begründeten Interpretationstheorie. Es ist vielmehr ein Sammelbegriff für alle literaturwissenschaftlichen Verfahren, bei denen die intensive Auseinandersetzung mit dem Text als Objekt im Mittelpunkt des Interesses steht. Vor allem soll der Begriff dem in den *Digital Humanities* – ebenso als Sammelbegriff – eingeführten Konzept des *Distant Reading* gegenübergestellt werden, der für Verfahren des indirekten, meist algorithmischen Zugangs zu Texten steht, vgl. Franco Moretti, „Conjectures on World Literature“, in: *New Left Review* 1 (2000), 54–68. Der Text als Mittelpunkt des Interesses ist beiden Zugängen gemein; der Unterschied liegt darin, wie stark der Fokus auf die konkrete Textoberfläche im Sinne ihrer Analyse bzw. Interpretation gerichtet ist.

⁴ „Algorithmic criticism seeks a new kind of audience for text analysis – one that is less concerned with fitness of method and the determination of interpretive boundaries, and one more concerned with evaluating the robustness of the discussion that a particular procedure announces“, siehe Stephen Ramsay, „Algorithmic Criticism“, in: Raymond G. Siemens/Susan Schreibman (Hg.), *A companion to digital literary studies*, Malden, MA, 2008, 477–491, hier: 490–491.

Humanities (TaDiRAH),⁵ die u. a. entwickelt wurde, um Vorhaben dabei zu unterstützen, „information on digital humanities tools, methods, projects, or readings“ zusammenzutragen, Annotation als erläuternd oder strukturell/linguistisch eingeführt und zur Forschungsaktivität der „Anreicherung“ („Enrichment“) gezählt.⁶ Da der Ansatz in TaDiRAH generisch ist, um möglichst umfassend sein zu können, könnte man Annotation im Sinne des *Close Reading* in der Literaturwissenschaft als immerhin mitgemeint auffassen. Allerdings ist TaDiRAH so angelegt, dass die Forschungsaktivitäten „Analysis“ und „Interpretation“ auf „Enrichment“ aufbauen, Annotation in diesen Aktivitäten aber keine Rolle spielt. Das ist, zusammen mit der Aussage, dass „die Taxonomie auch eine Modalität des Nachdenkens darüber [ist], was die digitalen Geisteswissenschaften sind“,⁷ aus Perspektive der literaturwissenschaftlichen Textanalyse unbefriedigend.

Andererseits gab es in den letzten zwanzig Jahren eine Reihe von Beiträgen zur Textanalyse, die als Fürsprecher für Annotation im Sinne des digitalen *Close Reading* gelesen werden können.⁸ Sie thematisieren oder exemplifizieren die

⁵Vgl. zu TaDiRAH z. B. Luise Borek/Christof Schöch/Klaus Thoden, „Digitale Forschungsaktivitäten multilingual: TaDiRAH für die deutschsprachige DH-Community“, in: *DHd 2016 Modellierung, Vernetzung, Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma. Konferenzabstracts*, Leipzig 2016, 376–378, sowie <http://tadirah.dariah.eu/vocab/sobre.php> (letzter Aufruf 15.09.2017).

⁶„Annotating refers to the activity of making information about a digital object explicit by adding, e.g., comments, metadata or keywords to a digitized representation or to an annotation file associated with it. This can be in the form of annotations that comment on or contextualize a passage (explanatory annotations) in order to make structural or linguistic information explicit (structural/linguistic annotation), as linked open data making the relationships between objects machine-readable, or, in the case of general metadata, adding information about the object as a whole. Encoding is a technique associated with annotating, as are POS-Tagging, Tree-Tagging, and Georeferencing“, <http://tadirah.dariah.eu/vocab/index.php?tema=22&annotating%20> (letzter Aufruf 15.9.2017).

⁷Borek/Schöch/Thoden (Anm. 5), 377.

⁸Z.B. Ian Lancashire, *Using TACT with electronic texts. A guide to text-analysis computing tools, version 2.1 for MS-DOS and PC DOS*, New York 1996; John Unsworth, *Scholarly Primitives. What methods do humanities researchers have in common, and how might our tools reflect this?*, London 2000, <http://people.brandeis.edu/~unsworth/Kings.5-00/primitives.html> (letzter Aufruf 15.9.2017); Lou Burnard, „On the Hermeneutic Implications of Text Encoding“, in: Domenico Fiormonte/Jonathan Usher (Hg.), *New Media in the Humanities. Research and Applications. Proceedings of the First Seminar „Computers, Literature, and Philology“*, Edinburgh, 7–9 September 1998, Oxford 2001, 31–38; Jerome McGann, „Marking Texts of Many Dimensions“, in: Susan Schreibman/Raymond G. Siemens/John Unsworth (Hg.), *A companion to digital humanities*, Malden, MA, 2004, 198–217; Allen H. Renear, „Text Encoding“, in: Susan Schreibman/Raymond G. Siemens/John Unsworth (Hg.), *A companion to digital humanities*, Malden, MA, 2004, 218–239; Wendell Piez, „Towards Hermeneutic Markup. An Architectural Outline“, in: *Digital Humanities 2010. Center for Computing in the Humanities, King's College. Conference Abstracts* (2010), 202–205, <http://piez.org/wendell/papers/dh2010/> (letzter Aufruf 15.9.2017).

interpretierende Auszeichnung von Texten als geisteswissenschaftliche Praxis⁹ und haben durch fortwährende Referenzierung z. T. einen nahezu kanonischen Status erlangt. Aber auch dies scheint keinen nachhaltigen Effekt zu haben. In Einführungswerken und Forschungspublikationen wird kaum auf manuelle Annotation als literatur- oder zumindest geisteswissenschaftliche Praxis Bezug genommen – teilweise selbst dann nicht, wenn auf die genannten Texte verwiesen wird.

So wird im Kapitel *Annotations* in der *Einführung in die Digital Humanities* von Jannidis/Kohle/Rehbein¹⁰ neben Unsworth und Renear sogar McGann zitiert.¹¹ Die Darstellung der Rolle von Annotation konzentriert sich aber auf Aspekte der Standardisierung, Disambiguierung und Edition. Zu Verfahren, die potenziell digitales *Close Reading* betreffen, findet sich nur ein Hinweis auf „private“ Annotationen.¹² Ebenso geht es an allen weiteren Stellen in der *Einführung*, in denen Annotation thematisiert wird, primär um ihre Rolle bei der Datenaufbereitung.

Auch im neuen, stark überarbeiteten *A New Companion to Digital Humanities*¹³ findet sich kaum etwas zu Annotation im Sinne des digitalen *Close Reading*. Der Index weist für Annotation nur einen Untereintrag zu virtuellen

⁹Lancashire (Anm. 8), 150–274, ergänzt seine Einführung in die Nutzung der Textanalysetools TACT mit Ausschnitten aus 40 literaturwissenschaftlichen Textanalysen, die mit dem Tool durchgeführt wurden.

Unsworth (Anm. 8) nennt „Annotating“ als eine von sieben grundlegenden geisteswissenschaftlichen Arbeitsweisen („primitives“).

Burnard (Anm. 8) stellt bei seinen Betrachtungen zu Markup fest: „The term ‚markup‘ covers a range of interpretive acts. [...] And markup is also able to represent characterizations such as analysis, interpretation, the affect of a text, or the contexts in which it was or is to be articulated“.

McGann (Anm. 8) verweist auf die textuellen Funktionen von Texten, die bei ihrer Auszeichnung berücksichtigt werden müssen, und entwickelt ein sechsdimensionales Textmodell. Nach der Anwendung an einem Gedicht von Robert Creeley schließt er: „This model of text-processing is open-ended, discontinuous, and non-hierarchical“ (212).

Renear (Anm. 8), 218, stellt seinem Beitrag zu *Text Encoding* folgendes Zitat von Michael Sperberg-McQueen voran: „Before they can be studied with the aid of machines, texts must be encoded in a machine-readable form. Methods for this transcription are called, generically, ‚text encoding schemes‘; such schemes must provide mechanisms for representing the characters of the text and its logical and physical structure ... ancillary information achieved by analysis or interpretation [may be also added]“ (Auslassungen/Ergänzungen im Original, E.G.).

Piez (Anm. 8) etabliert den Begriff des ‚hermeneutischen Markups‘: „By ‚hermeneutic markup‘ I mean markup that is deliberately interpretive. It is not limited to describing aspects or features of a text that can be formally defined and objectively verified. Instead, it is devoted to recording a scholar’s or analyst’s observations and conjectures in an open-ended way“.

¹⁰Andrea Rapp, „Manuelle und automatische Annotation“, in: Jannidis, Fotis/Kohle, Hubertus/Rehbein, Malte (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart 2017, 253–267.

¹¹Unsworth, Renear, McGann (jeweils Anm. 8).

¹²Rapp (Anm. 10), 260; vgl. dazu auch Fußnote 51.

¹³Susan Schreibman/Raymond G. Siemens/John Unsworth (Hg.), *A New Companion to Digital Humanities*, Chichester, West Sussex, UK, 2016.

Welten („virtual worlds and annotation“¹⁴) sowie den Eintrag „text markup“ aus, der – ebenso wie alle Untereinträge – auf McGanns bereits erwähnten, unveränderten Beitrag verweist. Alle weiteren Markup-bezogenen Verweise beziehen sich auf Markup-Sprachen und damit auf Standards, nicht auf Methoden.¹⁵ Von den sechs textorientierten Beiträgen im Abschnitt *Analysis* stammen zwei aus dem alten *A Companion to Digital Humanities*,¹⁶ die anderen vier sind neu. Abgesehen von McGanns Beitrag thematisiert aber kein Artikel die interpretatorische Praxis bzw. den Status von Annotation als literaturwissenschaftliche Tätigkeit.

Eine oberflächliche Suche in den einschlägigen, frei digital verfügbaren Publikationen bestätigt diesen Befund. Im *A Companion to Digital Humanities*, dem *Companion to Digital Literary Studies*¹⁷ sowie den Zeitschriften *Digital Scholarship in the Humanities* (und der Vorgängerin *Literary and Linguistic Computing*), *Digital Humanities Quarterly* und der *Zeitschrift für digitale Geisteswissenschaften* sind mittlerweile beinahe 1000 Artikel und Beiträge erschienen. Die Suche nach den Begriffen ‚Annotation‘ oder ‚Markup‘ sowie ihren jeweiligen Varianten ergibt immerhin fast 500 Treffer. Auch wenn einige Dopplungen dabei sein dürften, so scheint Annotation doch ein Kernkonzept im Bereich der *Digital Humanities* zu sein. Beschränkt man allerdings die Treffer auf jene, in denen außerdem ‚manual‘ oder ‚manuell‘/‚händisch‘ bzw. deren Varianten vorkommen, schrumpft die Menge auf etwas über 120 Beiträge. Von diesen dürfte wiederum ein Teil irrelevant sein, weil etwa der Begriff ‚manual‘ im Sinne eines Handbuchs verwendet wird oder die Suchbegriffe ausschließlich in den Literaturangaben vorkommen. Die genaue Auswertung der Treffer steht noch aus,¹⁸ aber eine kursorische Durchsicht der Texte bestätigt: Manuelle Annotation wird hauptsächlich im Rahmen von digitalen Editionen und von automatisierter Textanalyse thematisiert. Die Probleme der Textanalyse, die mit der manuellen Annotation im literaturwissenschaftlichen Bereich entstehen, werden hingegen kaum explizit adressiert.

Dies kann als Zeichen dafür gewertet werden, dass es sich bei manuellen Verfahren aus *Digital-Humanities*-Perspektive doch eher um einen (rein) literaturwissenschaftlichen Zugang handelt, der entsprechend wenig interessant ist.

¹⁴Ebd., 566.

¹⁵Die Verweise im Index sind: „Extensible Markup Language (XML)“, „markup languages“, „XML (Extensible Markup Language)“, siehe Schreibman/Siemens/Unsworth (Anm. 13), 557, 560 u. 566.

¹⁶Schreibman/Siemens/Unsworth (Anm. 8). Darin: neben McGanns Text (Anm. 8) der Beitrag von Michael Sperberg-McQueen: „Classification and its Structures“, 161–176.

¹⁷Susan Schreibman/Raymond G. Siemens (Hg.), *A companion to digital literary studies*, Malden, MA, 2008.

¹⁸Für die Auswertung muss außerdem ein weiterer Suchdurchgang durchgeführt werden, für den alternative Schlagwörter identifiziert werden. Dies ist sinnvoll, da sich bislang kein Begriff für Annotationen etabliert hat, der für die in diesem Beitrag ausgeführte Tätigkeit des digitalen *Close Reading* genutzt wird. Darüber hinaus sollte die Suche auch auf angrenzende Bereiche ausgeweitet werden, wie etwa die Literaturtheorie.

Genauso kann es als ein Befund interpretiert werden, der auf ein Versäumnis verweist. Plausibel sind beide Erklärungen. Eine Diskussion über die Ursachen dafür erscheint mir jedoch zum jetzigen Zeitpunkt wenig sinnvoll. Deshalb soll jetzt das Augenmerk auf konkrete Annotationsprojekte und ihren literaturwissenschaftlichen Beitrag gerichtet werden.

3 Zwei exemplarische Fälle computergestützter Annotation

3.1 Manuelle Annotation im Projekt *Erzählen über Konflikte*

Als erstes Beispiel für digitales *Close Reading* soll mein Dissertationsprojekt diskutiert werden, das ich von 2008 bis 2013 bearbeitet habe.¹⁹ Es verfolgte im Rahmen einer narratologischen Textanalyse einen manuellen, computergestützten Zugang. Die zugrunde liegende Forschungsfrage war, inwiefern narrative Strukturen eines Textes Aussagen über seine Konflikthaftigkeit zulassen. Am Beispiel von Erzählungen über Arbeitskonflikte, die als Interviews erhoben wurden, wurde deshalb eine möglichst umfassende narratologische Analyse vorgenommen, die darauf abzielte, eine möglicherweise distinktive narrative Konfiguration von Konflikterzählungen – im Kontrast zu Erzählungen ohne Konfliktbezug – herauszuarbeiten. Dafür wurden unter Rückgriff auf Lahn/Meister²⁰ bzw. Hühn u. a.²¹ etwa 400 narratologische Kategorien in einem Korpus von 39 Erzählungen untersucht, um anschließend anhand der Verteilung der Phänomene auf konflikthafte und konfliktlose Erzählungen die narrative Konfiguration dieser Erzählungstypen zu beschreiben.²²

Die Forschungsfrage entstand ursprünglich in einem nicht-digitalen Kontext. Sie war als Beitrag zur Methodik der Konfliktbearbeitung intendiert, der gleichzeitig eine Überprüfung der Anwendbarkeit narratologischer Konzepte beinhaltete: Da bei Konfliktbearbeitungsmethoden wie etwa der *Mediation* konfliktrelevante Aspekte z. T. unter Rückgriff auf narrative Strukturen beschrieben werden, war es naheliegend, eine narratologische Untersuchung von Konflikt Narrativen zu unternehmen, um nach weiteren narrativen Strukturen zu suchen, die in der Konfliktbearbeitung als Indikatoren oder als Analyseansatz genutzt werden können.

¹⁹ Evelyn Gius, *Erzählen über Konflikte. Ein Beitrag zur digitalen Narratologie*, Berlin 2015.

²⁰ Silke Lahn/Jan C. Meister, *Einführung in die Erzähltextanalyse*, Stuttgart/Weimar 2013.

²¹ Peter Hühn/John Pier/Wolf Schmid (Hg.), *the living handbook of narratology*, Hamburg, <http://www.lhn.uni-hamburg.de/> (letzter Aufruf 15.9.2017).

²² Die Analyse wird in Gius (Anm. 19) ausführlich dargestellt, die für die hier angestellten Betrachtungen der Effekte des computergestützten Vorgehens relevanten Punkte werden außerdem herausgestellt in: Dies., „Narration and Escalation. An Empirical Study of Conflict Narratives“, in: *Diegesis* 5/1 (2016), 4–25. Im Folgenden wird deshalb nur eine cursorische Beschreibung dieser Effekte unternommen, um anschließend auf die Frage des Beitrags zum klassischen literaturwissenschaftlichen Analyseprozess eingehen zu können.

Der Zuschnitt der Untersuchung war damit in der Narratologie ein Novum, das verschiedene Herausforderungen mit sich brachte, die sich sowohl methodologisch als auch theoretisch auswirkten. Im Laufe der Arbeit habe ich dann das zeitgleich entwickelte Annotations- und Analysetool CATMA²³ als Werkzeug für die Textannotation hinzugezogen und etwa 28.000 Textstellen annotiert.

Auch wenn dies ursprünglich nicht geplant war: Der digitale Zugang hat die Analyse in der durchgeführten Form erst ermöglicht. Die Auszeichnung derart zahlreicher Phänomene ist in einem analog vorliegenden Text kaum so umsetzbar, dass sie erkenntnisgenerierend sein kann, da es praktisch unmöglich ist, eine Übersicht über die Phänomene und ihre Verteilung zu behalten. Noch problematischer sind Mehrfachannotationen, bei denen sich teilweise acht, neun oder mehr Analysekatégorien an einer Textstelle überlagern. Auch die Auszeichnung von Phänomenen, die längere oder disjunkte Textpassagen umfassen, ist in der analogen Praxis schwer so umsetzbar, dass die entsprechenden Textstellen in ihrem Umfang oder ihrer Zugehörigkeit zueinander offensichtlich sind. Da die Analyse der Texte außerdem iterativ erfolgt, wäre ein Unternehmen mit einem solchen Umfang als analoges und damit nicht ohne Weiteres auszählbares und durchsuchbares Verfahren schlichtweg nicht praktikabel.

Die neuen Möglichkeiten erzeugten aber auch neue Probleme: Der induktive Zugang ohne Vorannahmen über die Beschaffenheit der narrativen Struktur und die narratologische Herangehensweise, die keine a-priori-Hierarchisierung ihrer Analysekonzepte vorsieht, boten beide keine Ansatzpunkte für eine Reihenfolge, in der die Phänomene identifiziert werden sollen. In der narratologischen Praxis ist das irrelevant, da narratologische Untersuchungen normalerweise nicht induktiv vorgehen, sondern bestimmte, für interessant befundene narratologische Aspekte von Texten, wie etwa Zeitstrukturen oder Figurencharakterisierung, umfangreicher untersuchen. Im Falle der Analyse der narrativen Strukturen von Konflikterzählungen war dies anders: Der induktive Zugang verlangte nach einer umsichtigen Sequenzierung der Analyseschritte in Bezug auf die analysierten Phänomene. Schließlich bedeutet jede weitere Iteration im hermeneutischen Zirkel eine Anreicherung und Veränderung des Vorverständnisses bzw. eine Weiterentwicklung des erlangten Textverständnisses und damit einen Schritt weg vom induktiven Zugang, der zuerst sammelt und dann analysiert bzw. interpretiert. Die über 400 narratologischen Katégorien, die in 14 Phänomenbereichen (etwa ‚Figuren‘, ‚Redewiedergabe‘, ‚Raum‘ etc.)²⁴ organisiert waren, mussten deshalb in

²³Vgl. Jan C. Meister/Marco Petris/Evelyn Gius u. a., *CATMA* (Version v5.2, 18. Juli 2018), <https://doi.org/10.5281/zenodo.1470119> bzw. www.catma.de (letzter Aufruf 15.9.2017).

²⁴Folgende Phänomenbereiche wurden genutzt: (1) Darstellung des Erzählers, (2) Erzählerinnenposition, (3) Erzählebenen, (4) Zeitpunkt des Erzählens, (5) Darstellung des Adressaten, (6) Erzählperspektive, (7) Präsentation von Rede und mentalen Prozessen, (8) Zeitrelationen zwischen Diskurs und Geschichte, (9) Wissensvermittlung und Informationsvergabe, (10) Erzählen über das Erzählen, (11) Handlung, (12) Figuren, (13) Aspekte des Raums und (14) Aspekte der zeitlichen Situierung. Die Einteilung in diese Phänomenbereiche folgt im Wesentlichen Lahn/Meister (Anm. 20), ebenso die Reihenfolge der Nennung.

eine Analysereihenfolge gebracht werden. Für diese wurde die möglichst geringe Auseinandersetzung mit dem Text als Maßstab genommen. Die Texte sollten entsprechend nicht über das für das jeweilige Phänomen betrachtete Maß hinaus analysiert werden, um den Einfluss von weiterem Ko(n)textwissen oder Interpretationshypothesen aus der vorhergehenden Betrachtung anderer Phänomene auf die jeweilige Analyse möglichst gering zu halten. Um dies zu erreichen, wurde eine Priorisierung von Analysekonzepten entwickelt, die auf einer Gewichtung der Kategorien in Bezug auf ihre Komplexität basierte. Unter Rückgriff auf (i) den Grad an Interpretationsabhängigkeit (im Sinne von semantischer Interpretation und Weltwissen), (ii) die Menge des für das Phänomen typischen Textumfangs und (iii) den für die Analyse benötigten zusätzlichen intratextuellen Kontext wurde deshalb ein Ranking der Komplexität der Analysekonzepte erarbeitet. Die resultierende Reihenfolge der Analyse beginnt bei Zeit und Raum und reicht – u. a. über Redewiedergabe, Erzählebenen und Figuren – vom zeitlichen Verhältnis zwischen *discours* und *histoire* (also Ordnung, Dauer und Frequenz) bis zu Perspektive und schließlich Informationsvergabe und umfasst damit das Spektrum von „interpretationsarm“/„nur Worte bzw. Wortgruppen betreffend“/„kaum kontextabhängig“ bis hin zu „interpretationsintensiv“/„den ganzen Text betreffend“/„vom Gesamtkontext abhängig“.²⁵

Ein weiterer Effekt des induktiven Zugangs war, dass aufgrund des dadurch entstehenden Umfangs der Arbeit eine theoretische Auseinandersetzung mit den narratologischen Analysekonzepten nicht möglich war und deshalb nicht der Analyse vorangestellt wurde. Die Auseinandersetzung mit den Kategorien erfolgte nur so weit, wie es nötig war, um sie für die Anwendung zu operationalisieren, also in Guidelines zu überführen, die bis zu einem gewissen Grad wiederholbare Analysen ermöglichen.

Allerdings sind im Zuge der Operationalisierung und Anwendung durchaus Probleme offensichtlich geworden, die für die narratologische Theorie von Bedeutung sind. Diejenigen, die im Hinblick auf die Anwendung gelöst werden konnten, wurden genauer betrachtet. So wurde etwa eine Reihe von Konzeptdefinitionen präzisiert, um Zweifelsfälle eindeutig annotieren zu können. Auf andere Probleme konnte nur verwiesen werden. So etwa auf die Zusammenhänge zwischen narratologischen Kategorien, die durch den breiten Zugang offensichtlich wurden. Diese sind in der theoretischen Debatte bislang unbemerkt geblieben, sollten aber im Sinne einer Systematisierung der narratologischen Konzepte weiter diskutiert werden.

Ein anderer Aspekt der Zusammenhänge wurde jedoch im Rahmen des Projekts bearbeitet: Für die konkrete Textanalyse folgt nämlich aus den offengelegten Zusammenhängen, dass einige Phänomene erst sinnvoll analysiert werden können,

²⁵ Gius (Anm. 19), 83–89.

nachdem andere Phänomene untersucht wurden. So sind etwa narrative Ebenen aufgrund der mit ihnen verbundenen Analyse der erzählten Welten als ontologische Konstrukte Voraussetzung für alle narratologischen Phänomene, die die *histoire*-Ebene – also die Ebenen des Geschehens, nicht die des Textes – einer Erzählung betreffen oder mit dieser in Verbindung stehen. Solche vorausgesetzten Phänomene müssen bei der narratologischen Analyse nach vorne gezogen werden, um die auf ihnen aufbauenden weiteren Phänomene analysieren zu können.

3.2 Kollaborative manuelle Annotation im Projekt *heureCLÉA*

Die narratologischen Kategorien, die für die Analyse von Konflikterzählungen entwickelt wurden, flossen anschließend in das Projekt *heureCLÉA*²⁶, das von 2013 bis 2016 bearbeitet wurde und einen in zwei Hinsichten erweiterten Skopus hatte: Zum einen war das Ziel des Projekts, die narratologische Analyse von Zeitphänomenen in narrativen Texten zu automatisieren. Dies führte nicht nur auf literaturwissenschaftlicher Seite zu einem veränderten – und innerhalb des Fachs neuen – Einsatz von Annotationen. Auch für die computerlinguistische Seite des Projekts bedeutete die Fragestellung Neuland, denn die betrachteten Genette'schen Kategorien wie Ordnung, Dauer und Frequenz²⁷ waren dezidiert narratologisch und damit literaturwissenschaftlich ausgerichtet. Sie bewegten sich deshalb sowohl in ihrer Komplexität als auch in ihrer Relevanz für die Geisteswissenschaften weit jenseits gängiger Sprachverarbeitungsansätze zur Textanalyse – wie etwa dem *Part-of-Speech-Tagging* oder auch der *Named Entity Recognition*. Das Projekt verfolgte durch diese Berücksichtigung disziplinärer Bedarfe einen originär geisteswissenschaftlich motivierten Zugang und schlug eine Brücke zwischen traditionellem *Close Reading* und digitalem *Distant Reading*.²⁸

²⁶Das „heure“ in *heureCLÉA* steht für ‚Heuristik‘ und bezieht sich auf die Entwicklung eines heuristischen Moduls, das für die Textanalyse eingesetzt werden kann. „CLÉA“ wiederum steht für den genutzten Zugang: ‚Collaborative Literary Exploration and Annotation‘. Der Akzent auf dem ‚É‘ schließlich soll darauf aufmerksam machen, dass die literaturwissenschaftliche Analyse nicht ausschließlich mit bestehenden Textanalyseverfahren durchgeführt werden kann, sondern teilweise eigene Zugänge nötig sind. So wie man eigentlich sprachunabhängige Verfahren, die anhand der englischen Sprache entwickelt wurden, teilweise nicht direkt auf andere Sprachen anwenden kann, weil sie schon an den Selbstverständlichkeiten der Sprachen bzw. der Zeichensysteme – wie etwa diakritischen Zeichen – scheitern.

²⁷Gérard Genette, *Die Erzählung*, München 1998.

²⁸Thomas Bögel/Michael Gertz/Evelyn Gius u. a., „Gleiche Textdaten, unterschiedliche Erkenntnisziele? Zum Potential vermeintlich widersprüchlicher Zugänge zu Textanalyse“, in: *Dhd2015 Von Daten zu Erkenntnissen. Book of Abstracts*, Graz 2015, 151–158. In diesem Beitrag werden hauptsächlich die literaturwissenschaftlichen Aspekte der Arbeit in *heureCLÉA* dargestellt. Für die Beiträge des Projekts im Bereich der *Digital Humanities* und der Computerlinguistik sei verwiesen auf: Thomas Bögel/Jannik Strötgen/Michael Gertz, „Computational Narratology. Extracting Tense Clusters from Narrative Texts“, in: *LREC Conference Proceedings*, 15. November 2014, http://www.lrec-conf.org/proceedings/lrec2014/pdf/280_Paper.pdf (letzter Auf-

Für die Erstellung des manuell annotierten Korpus als Grundlage für die Automatisierung wurde ein kollaborativer Zugang gewählt, wonach jeweils mindestens zwei Annotatoren und Annotatorinnen Guideline-orientiert ein Phänomen im selben Text annotierten, um intersubjektiv abgesicherte Analyseergebnisse zu erhalten. Außerdem sollte die gegenseitige Beeinflussung der Annotatoren und Annotatorinnen vorerst möglichst gering gehalten werden, damit diese ein jeweils eigenes Textverständnis entwickeln können, bevor sie sich mit anderen darüber austauschen. Diese Anforderungen führten zu einer Struktur für den Analyseprozess, die folgende Analyseschritte vorsah:

1. Einarbeitung in die narratologische(n) Kategorie(n)
2. Lesen der Primärtexte
3. Vertrautmachen mit den Guidelines²⁹
4. Annotieren der Texte
5. Diskussion über Probleme und Zweifelsfälle bei der Annotation; Festlegung einer Lösung
6. Überprüfung und ggf. Überarbeitung der Annotationen

Da auch im *heureCLÉA*-Projekt die Reihenfolge der Analyse an der steigenden Komplexität der genutzten Konzepte ausgerichtet war, mussten für alle folgenden Analysen von weiteren narratologischen Phänomenen außerdem die vorangegangenen Analysen der bereits bearbeiteten Phänomene berücksichtigt werden.³⁰ Die Auflistung der Schritte kann auch als eine Bewegung zwischen

ruf 15.9.2017); Thomas Bögel/Michael Gertz/Evelyn Gius u. a., „Collaborative Text Annotation Meets Machine Learning: heureCLÉA, a Digital Heuristic of Narrative“, in: *DHCommons Journal*, 1 (Juli 2015), <https://doi.org/10.5281/zenodo.324059>; Thomas Bögel/Evelyn Gius/Janina Jacke u. a., „Beyond Pragmatics. Disciplinary Profits of Interdisciplinary Approaches“, in: *DH2015 Global Digital Humanities. Conference Abstracts*, Sydney 2015; Thomas Bögel/Evelyn Gius/Janina Jacke u. a., „From Order to Order Switch. Mediating between Complexity and Reproducibility in the Context of Automated Literary Annotation“, in: *Digital Humanities 2016 Conference Abstracts*, Krakow 2016, 379–382; Evelyn Gius/Janina Jacke, „Informatik und Hermeneutik. Zum Mehrwert interdisziplinärer Textanalyse“, in: *Sonderband der Zeitschrift für digitale Geisteswissenschaften* 1 (2015), https://doi.org/10.17175/sb001_006; Dies., „Kollaboratives Annotieren literarischer Texte“, in: *DHd 2016. Modellierung, Vernetzung, Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma. Konferenzabstracts*, Leipzig 2016, 240–243; Janina Jacke/Jan C. Meister, „Pushing Back the Boundary of Interpretation. Concept, Practice and Relevance of a Digital Heuristic“, in: *Digital Humanities 2014 – Book of Abstracts*, Lausanne 2014, 264–266.

²⁹Vgl. Evelyn Gius/Janina Jacke, *Zur Annotation narratologischer Kategorien der Zeit. Guidelines zur Nutzung des CATMA-Tagsets*, Hamburg 2015, <http://heureclea.de/wp-content/uploads/2016/11/guidelinesV2.pdf> (letzter Aufruf 15.9.2017).

³⁰Genau genommen würde dies auch ohne die Gewichtung der Kategorien der Fall sein. Im Falle der gewichteten Reihenfolge steigt die Wahrscheinlichkeit weiter, dass die vorherigen Analysen die weiteren Analysen beeinflussen.

Vorwissen (1., 3., 5.) und Text (2., 4., 6.) im Sinne des hermeneutischen Zirkels betrachtet werden. Wenn man das Vorwissen um das narratologische Kategorieninventar, Guidelines und den Austausch über das Analysieren der Texte erweitert und zum Textganzen neben den Primärtexten auch die Annotationen sowie deren Überarbeitung zählt, modelliert die entwickelte Systematik den hermeneutischen Zirkel im Kontext der kollaborativen computergestützten Analyse.³¹

Die Behandlung von nicht übereinstimmenden oder auch widersprüchlichen Annotationen musste außerdem aufgrund der interdisziplinären Ausrichtung des Projekts im Spannungsfeld zwischen literaturwissenschaftlichem Interpretationspluralismus und informationswissenschaftlichen Konsistenzanforderungen erfolgen. Das führte zu einer weiteren Systematisierung des Annotationsprozesses. Es ging dabei darum, ein Modell zu entwickeln, mit dem sowohl der Polyvalenz von literarischen Texten Rechnung getragen wird als auch der Umfang der Operationalisierung und der Iterationen der Annotation in Abhängigkeit von den zur Verfügung stehenden Ressourcen methodisch transparent gestaltet werden kann.³² Dabei wurde neben einem System zur Bearbeitung von fehlerhaften Annotationen ein Zugang entwickelt, der einen flexiblen Umgang mit den Fällen erlaubt, in denen divergierende Annotationen durch divergierende Voranalysen bzw. divergierende Verständnisse weiterer – mit den betrachteten Kategorien im Zusammenhang stehender – Kategorien bedingt sind. Neben der Sensibilisierung für derartige Fälle, die oft implizit sind, hat das Modell den Vorteil, dass der Umgang mit diesen Fällen dem Bedarf innerhalb des Forschungsprojekts angepasst werden kann, der sich aus der Forschungsfrage und den Rahmenbedingungen des Projekts zusammensetzt. So erzeugte die Bestimmung von Anachronien – also Vor- und Rückgriffen in der Geschichte (Prolepsen, Analepsen) – im *heureCLÉA*-Projekt einen unerwartet geringen Konsens zwischen den Annotatoren und Annotatorinnen. In der Diskussion über die Annotationen stellte sich heraus, dass die Annotatoren und Annotatorinnen unterschiedlich mit den Erzählebenen in den Texten umgegangen waren und entsprechend in den weniger umsichtig analysierten Fällen Erzählebenen mit Anachronien verwechselt wurden. Deshalb musste die schon im Vorgängerprojekt festgestellte Abhängigkeit zwischen Erzählebenen und anderen Phänomenen mit Bezug zur *histoire*-Ebene der Erzählung berücksichtigt werden. Entsprechend wurden die Erzählebenen der Texte annotiert. Anschließend wurde die Analyse der Anachronien erneut vorgenommen. Der entstandene Mehraufwand war erheblich: Die Erzählebenen mussten analysiert werden – ein aufgrund der Omnipräsenz des Phänomens in narrativen Texten recht aufwendiges Unterfangen. Anschließend musste eine Überprüfung und in vielen Fällen eine Neubearbeitung der Anachronie-

³¹Vgl. dazu auch die Abbildung des erweiterten hermeneutischen Zirkels in Evelyn Gius/Janina Jacke, „The Hermeneutic Profit of Annotation. On preventing and fostering disagreement in literary text analysis“, in: *International Journal of Humanities and Arts Computing* 11/2 (2017), 233–254, hier: 242, <https://doi.org/10.3366/ijhac.2017.0194>.

³²Vgl. für eine ausführliche Darstellung des Modells ebd., 241–250.

Annotationen erfolgen. Dies war nötig, um für die Automatisierung eine höhere Übereinstimmung zwischen den Annotatoren und Annotatorinnen zu erreichen, und möglich, weil die zusätzlichen Annotationen im Sinne einer narratologisch genauen Analyse waren. Ein anderer Zusammenhang, der ebenfalls narratologisch von Belang ist, wurde hingegen nicht umgesetzt: Bei der Analyse der Dauer – also des Verhältnisses zwischen dem zeitlichen Umfang der erzählten Geschichte und dem Umfang des Erzähltexts – waren die Annotationen z. T. widersprüchlich. Dieselbe Passage wurde von einer bestimmten Annotatorin als ‚gerafft erzählt‘ und von einem anderen Annotator als ‚gedehnt erzählt‘ annotiert. Es wurde also einmal behauptet, der Text der Passage sei im Verhältnis zum durch ihn erzählten Geschehen kurz, und einmal, der Text sei verhältnismäßig länger. Diese Widersprüche ließen sich auf das zugrunde liegende implizite Verständnis von Ereignishaftigkeit zurückführen. Konkret ging es um die Frage, ob im betreffenden Textteil etwas geschieht, also Ereignisse im narratologischen Sinne vorliegen, oder die Passage nicht narrativ ist.³³ Allerdings ist das Ereigniskonzept selbst in der Narratologie umstritten, und außerdem ist nicht absehbar, welche weiteren Interdependenzen zwischen Ereignissen und anderen narratologischen Kategorien bestehen. Gleichzeitig mussten wir aber von einer Interferenz mit den Zeitphänomenen ausgehen, da Ereignisse im narratologischen Verständnis Zustandsveränderungen sind und damit immer eine Zeitdimension besitzen. Neben diesen narratologischen Vorbehalten gab es noch einen pragmatischen: Da Ereignisse als „smallest indivisible unit“ narrativer Texte gelten,³⁴ würde die Annotation von Ereignissen einen kaum vertretbaren theoretischen wie praktischen Aufwand bedeuten. Deshalb wurde in diesem Fall ein vom Modell vorgesehener Mittelweg gewählt: Die Annotationen der Dauerphänomene wurden von allen Annotatoren und Annotatorinnen um Begründungen ergänzt, die das ihrer Analyse zugrunde liegende Verständnis von Ereignishaftigkeit explizieren. Gleichzeitig wurden alle dazu angehalten, darauf zu achten, dass sie ihr Konzept von ‚Ereignishaftigkeit‘ konsistent anwenden.

Die Auseinandersetzung mit den Fällen der Nichtübereinstimmung von Annotationen, die auf Zusammenhänge mit weiteren narratologischen Kategorien zurückzuführen sind, brachte auch systematische Zusammenhänge zwischen den bekannten narratologischen Kategorien der Zeit (Ordnung, Dauer, Frequenz) und narrativen Ebenen ans Licht. So ist, wie oben für die Analyse der Anachronien (Ordnung) erläutert, im Fall der Analyse von Ordnungs- und Dauerphänomenen ein mehrschrittiges Verfahren notwendig. Dabei wird jede narrative Ebene zunächst

³³Neben der Gegenüberstellung von deskriptiven und narrativen Textelementen spielt hier auch die Frage der Bedingungen für Ereignisse bzw. Ereignishaftigkeit eine Rolle. So wird die Realität von Zustandsveränderungen (in der fiktiven Welt) u. a. von Wolf Schmid (*Elemente der Narratologie*, Berlin/Boston 2008, 21) als Bedingung für Ereignishaftigkeit genannt und durch die zusätzliche Bedingung der Resultativität, also das zum Abschluss gekommene Sein der Veränderung, verstärkt.

³⁴Vgl. Jurij Lotman, *The Structure of the Artistic Text*, Ann Arbor 1977, 233.

einzelnen und erst anschließend unter Berücksichtigung weiterer – ggf. eingebetteter – Ebenen analysiert. Im Falle von Ordnung und Frequenz ist wiederum eine ontologische Bestimmung der analysierten Ebenen notwendig, da nur jene Ereignisse Kandidatinnen für die Ordnungs- und Frequenzphänomene sind, die auf derselben ontologischen Ebene, also in derselben fiktionalen Welt, stattfinden.

4 Computergestütztes *Close Reading* als literaturwissenschaftliches Forschungsparadigma?

4.1 Zur hermeneutischen Textanalyse

Inwiefern entsprechen nun die beschriebenen digitalen Arbeiten dem literaturwissenschaftlichen Paradigma der ‚hermeneutischen Textanalyse‘? Beide Projekte zeichnen sich zweifelsohne durch eine intensive Auseinandersetzung mit den Primärtexten aus und haben dementsprechend Texte als Objekte im Fokus. In den Projekten wurden u. a. etwa 28.000 bzw. 32.000 Textstellen annotiert, wobei die analysierten Primärtexte jeweils nur etwa 60.000 bzw. 80.000 Tokens umfassten. Wenn man das auf DIN-A4-Seiten umrechnet, sind das, abhängig von der angesetzten Zeichenzahl einer Seite, 100 bis 150 Seiten. Damit finden sich selbst bei einer vorsichtigen Schätzung von 150 Seiten durchschnittlich über 400 Annotationen auf einer Seite.³⁵

Um zu untersuchen, inwiefern die Arbeiten über das *Close Reading* hinaus einen Bezug zur literaturwissenschaftlichen bzw. hermeneutischen Textanalyse haben, muss diese genauer beschrieben werden. Hier kommt allerdings das bereits oben im Kontext der vorgenommenen Kontrastierung angesprochene Problem zum Tragen: Die hermeneutische Textanalyse ist nicht nur in konkreten Anwendungsfällen kaum methodisch expliziert, es gibt auch generell keine einheitliche Auffassung über das Verfahren. Der Grund dafür liegt in der Vielfalt literaturwissenschaftlicher Methoden, die „nicht in Reinform vorkommen, sondern Bestandteil einer Interpretationstheorie sind“, wobei zudem „statt genauer Angaben über einzelne Interpretationsschritte meist unspezifische Regeln

³⁵Die Korpora und Annotationen sind in folgenden Publikationen im CATMA-generierten XML-Format veröffentlicht und nachnutzbar: Evelyn Gius, *Korpus „Erzählen über Konflikte“* (August 2013), <https://doi.org/10.5281/zenodo.894732>; Dies., *Narratologische Annotationen des Korpus’ „Erzählen über Konflikte“* (August 2013), <https://doi.org/10.5281/zenodo.894720>; Evelyn Gius/Janina Jacke/Jan C. Meister u. a., *Heurecléa Source Documents: 1.0.* (2017), <https://doi.org/10.5281/zenodo.274962>; Dies., *Heurecléa Time Annotations Compared Public: VI.1.* (22. Februar 2017), <https://doi.org/10.5281/zenodo.595484>; Dies., *Heurecléa Time Annotations Compared Public: VI.1.* (22. Februar 2017), <https://doi.org/10.5281/zenodo.595484>; Dies., *Heurecléa Time Annotations Uncompared Public: VI.1.* (22. Februar 2017), <https://doi.org/10.5281/zenodo.593119>.

formuliert werden“.³⁶ Versuche einer einheitlichen Beschreibung der literaturwissenschaftlichen Textanalyse gibt es deshalb nur in Einführungswerken oder Fachlexika. Die Beschreibungen in Einführungen sind – eher problematisierend als explizierend – darauf ausgerichtet, bei den Studierenden ein Verständnis für die Diversität der Zugänge in der Literaturwissenschaft zu entwickeln, das sich gleichzeitig von nicht-wissenschaftlichen Verfahren abgrenzt. Anstelle von detaillierten Beschreibungen wird das Vorgehen beispielhaft skizziert und mit allgemeinen Feststellungen oder Hinweisen ergänzt.³⁷ Einführungen können also keine Grundlage für die Darstellung literaturwissenschaftlicher Textanalyse sein. Deshalb basiert das in diesem Beitrag genutzte Verständnis des hermeneutischen Verfahrens auf einer Zusammenführung aus den für die Textanalyse und -interpretation einschlägigen Artikeln im *Reallexikon der deutschen Literaturwissenschaft* zu den Lemmata ‚Hermeneutik‘, ‚Hermeneutischer Zirkel‘, ‚Interpretation‘ und ‚Textanalyse‘.³⁸

Wie bereits oben eingeführt, wird ‚Hermeneutik‘ in Anlehnung an den gleichnamigen Eintrag im *Reallexikon* als auf das Lesen, Verstehen und Interpretieren von Texten ausgerichtet verstanden und nicht als die Umsetzung eines konkreten hermeneutischen Verfahrens im Sinne Schleiermachers, Gadamers oder anderer Theoretiker und Theoretikerinnen. Vielmehr werden mit dem

³⁶ Simone Winko, „Methode, literaturwissenschaftliche“, in: Ansgar Nünning (Hg.), *Metzler-Lexikon Literatur- und Kulturtheorie*, Stuttgart 2008, 495–496, hier: 495.

³⁷ So erläutert etwa Jochen Vogt in seiner *Einladung zur Literaturwissenschaft* zur Hermeneutik der Entfaltung, die er als (die) literaturwissenschaftliche Methode der Textanalyse und -interpretation einführt: „Sie wird einerseits, den hermeneutischen Regeln von Schleiermacher bis Habermas folgend, den **Sachgehalt** ihrer Texte im *Kommentar erschließen*. Sie muss aber andererseits immer die **prinzipielle Vieldeutigkeit literarischer Texte** bedenken und aus ihren jeweiligen Kommentaren nur **behutsame Deutungen** ableiten. Jede einzelne muss sich dabei um **Plausibilität und Nachvollziehbarkeit bemühen**, d. h. möglichst viele Aspekte eines gegebenen Textes integrieren und dabei anderen, konkurrierenden Deutungen standhalten können. Vor allem aber wird diese literarische Hermeneutik darauf achten müssen, durch welche Kunstmittel ein Text Bedeutung, besser: Bedeutungen, Sinn-Potenziale erzeugt. Szondi und Ricoeur haben, wie lange vor ihnen schon Schleiermacher, in diesem Sinn vor allem die Interpretation der *Metaphorik*, also des bildlichen Sprechens, als ein Kernstück literarischer Hermeneutik herausgestellt.“ Aus: Jochen Vogt 2008, *Einladung zur Literaturwissenschaft. Mit einem Vertiefungsprogramm im Internet*, München/Paderborn ⁶2008, 69, Fettungen von E.G.; diese Beschreibung ist ein kleinster gemeinsamer Nenner, in dem sich eine Reihe an Interpretationstheorien wiederfindet – und deshalb zu unspezifisch, um für eine Untersuchung konkreter methodischer Schritte genutzt zu werden.

³⁸ Weimar (Anm. 1); Klaus Weimar, „Hermeneutischer Zirkel“, in: Harald Fricke/Klaus Grubmüller/Jan-Dirk Müller u. a. (Hg.), *Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*, Berlin 2000, 31–33; Axel Spree, „Interpretation“, in: Harald Fricke/Klaus Grubmüller/Jan-Dirk Müller u. a. (Hg.), *Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*, Berlin 2000, 168–172; Simone Winko, „Textanalyse“, in: Harald Fricke/Klaus Grubmüller/Jan-Dirk Müller (Hg.), *Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*, Berlin 2003 597–601.

Begriff ‚hermeneutisch‘ zwei Aspekte herausgestellt, die wesentlich sind für literaturwissenschaftliche Textanalysepraxis und die Reflexion darüber: Der hermeneutische Zugang zu Texten bedeutet zum einen, dass das Verstehen – im Gegensatz zum Erklären – im Zentrum der Aktivitäten steht. Entsprechend zielen weiterhin die Bemühungen, die unter ‚Hermeneutik‘ gefasst werden können, auf die Rekonstruktion der Phasen und des Verlaufs des Textverstehens in einem Verstehensmodell ab.³⁹ Die daraus resultierenden Modelle können unter den sogenannten ‚hermeneutischen Zirkel‘ gefasst werden, der wiederum eine Beschreibung für die „[w]echselseitige Abhängigkeit von Verständnis eines Ganzen und seiner Teile“ ist.⁴⁰ Es geht beim Textverstehen entsprechend darum, diese Abhängigkeit, die sich als Dilemma darstellt, aufzulösen. Dies geschieht durch das sogenannte Vorverständnis des Textganzen oder seiner Teile, das das Verstehen leitet, wobei es vom Text bestätigt oder revidiert werden kann.⁴¹ Führt man dieses Vorgehen zurück auf die zentralen Aktivitäten der Hermeneutik – das Lesen, Verstehen und Interpretieren von Texten –, so ermöglicht die Unterscheidung zwischen Textanalyse und Interpretation die differenziertere Beschreibung literaturwissenschaftlicher Tätigkeiten.⁴² Unter ‚Textanalyse‘ werden wiederum zwei Komponenten gefasst: die Beschreibung und die Analyse. Die Beschreibung beinhaltet zumeist rein deskriptive Verfahren, die auf die quantifizierbaren Faktoren sowie sprachlichen Oberflächenphänomene eines literarischen Textes abzielen. Die Analyse zielt hingegen auf Strukturbildung und untersucht entsprechend „globalere strukturelle Merkmale eines Textes“ wie *Isotopien* oder andere interne semantische Beziehungen zwischen Textelementen.⁴³

‚Textanalyse‘ – also Beschreibung und Analyse – ist außerdem die „Vorstufe und Bedingung der Interpretation“, wobei die Grenzen zwischen der Analyse und der Interpretation von Texten aus erkenntnistheoretischer Sicht fließend sind.⁴⁴ Grundsätzlich sind aber Bedeutungszuweisungen im Rahmen der Textanalyse

³⁹Weimar (Anm. 1), 28.

⁴⁰Weimar (Anm. 39), 31.

⁴¹Ebd., 32.

⁴²Winko (Anm. 39), 597. Dies gilt insbesondere für textimmanente oder textorientierte Zugänge wie die in diesem Beitrag vorgestellten narratologischen Analysen. Vgl. dazu Tom Kindt/Hans-Harald Müller, „Zum Verhältnis von Interpretation und Deskription. Ein Bestimmungsvorschlag und ein Beispiel“, in: Jan Borkowski/Stefan Descher/Felicitas Ferder/Philipp D. Heine (Hg.), *Literatur interpretieren. Interdisziplinäre Beiträge zur Theorie und Praxis*, Münster 2015, 73–90; Janina Jacke, „Is There a Context-Free Way of Understanding Texts? The Case of Structuralist Narratology“, in: *Journal of Literary Theory* 8/1 (2014), 118–39, <https://doi.org/10.1515/jlt-2014-0005>.

⁴³Winko (Anm. 39), 598.

⁴⁴Ebd., 597 f.

elementarer bzw. textnäher als bei einer Interpretation und damit leichter überprüfbar.⁴⁵

Die Unterscheidung zwischen den Begriffen ‚Textanalyse‘ und ‚Interpretation‘ scheint teilweise analog zur Unterscheidung zwischen ‚Textteilen‘ und ‚Textganzen‘ zu sein und entsprechend die beschriebenen Abhängigkeiten im hermeneutischen Zirkel abzubilden. In beiden Fällen sind allerdings keine trennscharfen Unterscheidungen der konstituierenden Elemente möglich, was die Grenzen zwischen den Begriffen ‚Analyse‘ und ‚Interpretation‘ ebenso fließend macht wie jene zwischen ‚Verständnis von Textteilen‘ und ‚Verständnis des Textganzen‘ oder jene zwischen ‚Erklären‘ und ‚Verstehen‘.

Trotz dieser Unschärfen soll für die folgenden Betrachtungen ‚hermeneutisches Textverstehen‘ definiert werden als dreistufige Auseinandersetzung mit Texten in Form von Beschreibung, Analyse und Interpretation. Diese Auseinandersetzung findet außerdem iterativ statt, wobei die Iterativität durch zwei Typen von Abhängigkeiten bedingt ist: Zum einen werden alle drei Aktivitäten (Beschreibung, Analyse und Interpretation) durch ein heuristisches Vorverständnis beeinflusst, zum anderen greift die jeweils höherkomplexe Aktivität auf die weniger komplexe(n) Aktivität(en) zurück.

4.2 Computergestütztes *Close Reading* als literaturwissenschaftliches Verfahren

Das entwickelte Verständnis hermeneutischen Textverstehens soll nun als Grundlage für eine Bewertung der beschriebenen Aktivitäten und Entwicklungen in den beiden Projekten dienen. Entsprechend wird im Folgenden dargelegt, welche Aspekte der Forschungsarbeit unter die drei wesentlichen Aktivitäten – Beschreibung, Analyse und Interpretation – der hermeneutischen Textanalyse fallen und welche das quer zu diesen verlaufende Vorverständnis betreffen. Der Zusammenhang zwischen Beschreibung, Analyse und Interpretation, der durch den Aufbau der jeweils komplexeren Tätigkeiten auf den weniger komplexen Aktivitäten begründet ist, wird nicht eigens adressiert, da die Abgrenzung der drei Tätigkeiten nicht ohne Weiteres möglich ist. Er ist aber z. T. in den jeweiligen Darstellungen und insbesondere in den übergreifenden Ausführungen zum Vorverständnis erkennbar.

Noch ein Hinweis zur Verwendung des Terminus ‚Analyse‘: Im *Reallexikon* werden unter Verweis auf die gängigen Verwendungsweisen des Begriffs bzw. daraus zusammengesetzter Begriffe mehrere Tätigkeiten der literaturwissenschaftlichen

⁴⁵Ebd., 598. Außerdem konzentrieren sich Bedeutungszuweisungen im Rahmen der Textanalyse auf intratextuelle Kontexte und ziehen nur jene extratextuellen Kontexte heran, die für ein primäres Verständnis notwendig sind. Die Ergebnisse der Textanalyse dienen wiederum als Grundlage zur Interpretation, wobei inter- und extratextuelle Kontexte hinzugezogen werden, um Hypothesen über die Bedeutung des Textes als Ganzes zu generieren.

Auseinandersetzung mit Texten erläutert, von denen drei für diesen Beitrag relevant sind. Da die Einführung weiterer, neuer Begriffe für diese Tätigkeiten nicht ohne Weiteres möglich ist und potenziell andere Arten von Verwirrung verursacht, werden die Begriffe aus dem *Realllexikon* im Wesentlichen übernommen. ‚Textanalyse‘ wird also als Oberbegriff für die hauptsächlich intratextuell ausgerichteten Tätigkeiten genutzt, die durch diese Ausrichtung von der Interpretation abgegrenzt werden können. ‚Analyse‘ bezeichnet eine der beiden intratextuellen, textanalytischen Tätigkeiten. Sie kann, da sie komplexer auf die Textstruktur ausgerichtet ist, von der anderen Komponente, der ‚Beschreibung‘, abgegrenzt werden. Die im *Realllexikon* ebenfalls unter dem Terminus ‚Textanalyse‘ gefasste, mit der Untersuchung literarischer Texte zusammenhängende Tätigkeit wird schließlich als ‚hermeneutische Textanalyse‘ bezeichnet. Damit werden also sowohl die Analyse und Beschreibung umfassende Textanalyse als auch die Interpretation von Texten gefasst.

4.2.1 Beschreibung

„[D]ie überwiegend deskriptiven Verfahren, die sich auf Sprache und Form eines literarischen Textes richten“, gelten als ‚Beschreibung‘.⁴⁶ Im Falle der klassischen Narratologie ist zumindest ein Teil der Phänomene eher auf der Ebene des Textes bzw. der Textoberfläche anzusiedeln, deshalb kann ihre Bestimmung als ‚Beschreibung‘ gelten. Die hier diskutierten Projekte hatten außerdem eine nachfolgende Automatisierung der Analysen im Blick, weshalb eine möglichst textinterne Analyse angestrebt wurde und insbesondere Oberflächenphänomene eine Rolle in der Analyse spielen sollten. Da zudem selbst jener Teil der Textanalyse, der der Beschreibung dient, „technisch gesehen [...] auf interpretierenden Annahmen beruhen“ kann⁴⁷ und damit Elemente der Analyse aufweist, kann die narratologische Analyse zumindest auch als Textbeschreibung aufgefasst werden.

Wenn man davon ausgeht, dass die verwendete Terminologie für die Beschreibung von Texten zentral ist, fallen einige Aspekte der diskutierten Projekte unter die ‚Textbeschreibung‘. Beide Projekte haben unterspezifizierte Definitionen der narratologischen Konzepte offengelegt und ergänzt. Die im Laufe der Begriffsarbeit außerdem entdeckten Abhängigkeiten zwischen narratologischen Analysekategorien tragen ebenfalls zu diesem Aspekt der hermeneutischen Textanalyse bei. Wenn sich die Abhängigkeiten als stabil herausstellen, müssten die narratologischen Analysekategorien entsprechend überarbeitet werden. Neben der Systematisierung des Begriffsapparats im Sinne offengelegter Abhängigkeiten wäre dadurch außerdem eine Redundanzreduktion zu erwarten.⁴⁸ Die damit weiterentwickelte Terminologie stellt eine verbesserte Struktur für die Untersuchung literarischer Texte im Sinne einer Beschreibung dar.

⁴⁶ Ebd.

⁴⁷ Ebd.

⁴⁸ Gius (Anm. 19), 355.

4.2.2 Analyse

Bei der aus Beschreibung und Analyse bestehenden ‚Textanalyse‘ geht es hauptsächlich um Verfahren, die eine nachfolgende Interpretation des Textes strukturiert ermöglichen. Textanalyse ist die „Vorstufe und Bedingung der Interpretation“ bzw. ein „Verfahren zur Sicherung der Deutungsarbeit“.⁴⁹ Die ‚Analyse‘ als eine Komponente der Textanalyse betrifft im Vergleich zur anderen, bereits diskutierten Komponente, der ‚Beschreibung‘, komplexere Textphänomene. Je nach Auffassung über die Komplexität von narratologischen Kategorien kann auch die bereits diskutierte Arbeit an den Analysekatégorien in den Bereich der Analyse fallen. Daneben sind es insbesondere die konkreten Schritte, die in den beschriebenen Projekten beim Annotieren gemacht wurden, die als Beitrag im Bereich der ‚Analyse‘ gesehen werden können. In beiden Projekten entstanden im Forschungsprozess Anforderungen, die dazu führten, dass die Analyseroutine expliziert, erweitert oder überhaupt erst ausgearbeitet werden musste.

So etwa die Überlegungen zur Gewichtung der Analysekatégorien. Diese waren von der verfolgten Fragestellung zur narrativen Struktur von Konflikt-erzählungen unabhängig und sind deshalb generell auf narratologische Analysen übertragbar, ebenso wie die daraus resultierende Reihenfolge der Analyse. Beide gelten außerdem unabhängig vom digitalen Zugang. Damit führte der durch die Digitalität des Zugangs erst entstandene Bedarf eines neuen Vorgehens zu einer methodologischen Erweiterung der narratologischen Analysepraxis, die auch für die nicht-digitale Narratologie relevant ist und dort angewendet werden kann.

Des Weiteren kann die Integration der Annotationen in den hermeneutischen Prozess als ein die Analyse betreffender Aspekt betrachtet werden. Die Tatsache, dass die Annotationen teilweise Grundlage für darauf aufbauende Analysen sind und gewissermaßen als zusätzlicher Text für die anschließende Interpretation vorliegen, führt ein neues, distinktes Element in den Analyseprozess ein.⁵⁰

4.2.3 Interpretation

Die ‚Interpretation‘ der analysierten Texte, also die Frage, inwiefern die Ergebnisse der Textanalyse als Interpretationsgrundlage genutzt wurden, um Hypothesen über die Bedeutung des Textes als Ganzes zu generieren,⁵¹ ist ein wesentlicher Aspekt der meisten literaturwissenschaftlichen Praktiken im Umgang mit Texten. Allerdings wurde in der Beschreibung der beiden Projekte die

⁴⁹Winko (Anm. 39), 597.

⁵⁰In der Editionsphilologie werden Annotationen zwar als eigene, zum Text gehörige Paratexte betrachtet. So spricht etwa Rapp in ihrem Beitrag zu manuellen und automatischen Annotationen entsprechend von einem „eigenen Textsortentyp“. Allerdings wird die eben herausgestellte Möglichkeit, diesen Textsortentyp auch als Zugang zur Polyvalenz und Ambiguität von literarischen Texten einzusetzen und damit für eine Interpretation nutzbar zu machen, durch die Forderung nach einem „Höchstmaß an Einheitlichkeit und Vergleichbarkeit“ der Annotationen faktisch ausgeschlossen. Vgl. Rapp (Anm. 10), 257.

⁵¹Winko (Anm. 39), 598.

Interpretation der Texte praktisch nicht erwähnt. Das liegt nicht nur daran, dass ‚Interpretation‘ im Sinne einer ‚Hypothese über Textbedeutung‘ in der klassischen Narratologie kaum eine Rolle spielt.⁵² Die Bedeutung der Texte als Ganze spielte in diesen beiden Projekten eine besonders geringe Rolle, da es um Erkenntnisse über narrative Strukturen bzw. um die Erkennung dieser Strukturen ging.

Diese aus den Forschungsfragen heraus begründete Lücke im textanalytischen Prozess sollte aber nicht darüber hinwegtäuschen, dass im Kontext des digitalen *Close Reading* die Interpretation zumeist nicht nahtlos an die vorhergehenden Analysen anschließt. Das ist u. a. damit erklärbar, dass die meisten Tools zur Erstellung digitaler Annotationen nur – wenn überhaupt – eine Analyse der Texte ermöglichen, aber keine weitere Arbeit in Richtung Interpretation innerhalb des Tools unterstützt wird. Auch Tools wie CATMA, die explizit für die literaturwissenschaftliche Textanalyse entwickelt wurden und den iterativen Analyseprozess durch die nahtlose Integration von Annotations- und Suchfunktionalitäten modellieren, stellen ihren Nutzerinnen und Nutzern keine Funktionalitäten zur Verfügung, die diese zur Erarbeitung einer Interpretation nutzen können. Es gibt zwar seit einiger Zeit Ansätze, die die geisteswissenschaftlichen Forschungs- und Verstehensprozesse in einer digitalen Umgebung unterstützen, indem sie das Generieren von Hypothesen und sogar das Publizieren der entstehenden Interpretationen ermöglichen.⁵³ Doch während in den Sozialwissenschaften Werkzeuge wie NVivo, atlas.ti oder MAXQDA zur Verfügung stehen, die den gesamten Forschungsprozess umfassen, der auch dort im Wesentlichen vom Etablieren von Analysekatégorien über die Annotation und Analyse der Daten bis hin zur Entwicklung einer Interpretation⁵⁴ reicht, hat sich bislang in den textbasierten Geisteswissenschaften kein System etablieren können, das die bestehende Lücke zwischen der Analyse und der darauf aufbauenden Interpretation schließt. Dies mag daran liegen, dass in den Sozialwissenschaften der

⁵²Für eine differenzierte Diskussion der Interpretationsabhängigkeit narratologischer Zugänge vgl. Jacke (Anm. 43).

⁵³So macht etwa das Tool ‚Pliny‘ neben dem Annotieren verschiedener Primärquellen auch die Entwicklung von Argumenten und Interpretationen aus diesen sowie aus wiederverwendbaren Notizen möglich, die visuell organisiert und weiter strukturiert werden können, indem vielfältige Arten von Beziehungen zwischen ihnen ausgedrückt werden können, vgl. John Bradley, „Thinking about Interpretation. Pliny and Scholarship in the Humanities“, in: *Literary and Linguistic Computing* 23/3 (2008), 263–279, <https://doi.org/10.1093/lilc/fqn021>. In eine ähnliche, wenn auch lineare Richtung gehen ‚Notebooks‘, die das Schreiben, Programmieren und Ausführen von Code in derselben, publizierbaren Umgebung ermöglichen. Hierzu zählen z. B. das *iPython Notebook* von Jupyter (<https://ipython.readthedocs.io/en/stable/index.html>) und das sich noch in Entwicklung befindenden *Spiral Notebook*, das u. a. die Textanalysetools der Plattform *Voyant* integriert (<https://voyant-tools.org/spiral/alta> [jeweils letzter Aufruf 15.9.2017]).

⁵⁴Dafür werden u. a. Mindmap-artige Funktionen zur Verfügung gestellt, in denen Annotationen, Kategorien (*Codes*), Primärdaten, Analyseergebnisse, Notizen (*Memos*) und weitere hinzufügbare Elemente verbunden werden können. Die aus den Programmen stammenden Elemente sind direkt aus diesen Darstellungen heraus zugänglich, außerdem wird das iterative Arbeiten durch eine geeignete Versionierung unterstützt.

Ausdifferenzierungsprozess in ‚qualitative‘ und ‚quantitative Sozialwissenschaften‘ Jahrzehnte zurückliegt und entsprechend sowohl die für die Forschungsprozesse unterstützenden Tools schon lange bekannt sind als auch die Kontroverse über den geeigneten – qualitativen oder quantitativen – Zugang zum Forschungsmaterial mittlerweile eher unaufgeregt und dadurch zielorientiert geführt wird. Dadurch sind die Forschungsprozesse wesentlich klarer und die Ansätze zu ihrer Unterstützung vielfach erprobt. Diese Entwicklung zeichnet sich in den Geistes- und insbesondere den Literaturwissenschaften erst ab. Der Prozess der Interpretation muss dort noch genauer erforscht und beschrieben werden, um Tools entwickeln zu können, die wirklich geisteswissenschaftliche Analysepraktiken im Sinne des mittlerweile 50 Jahre alten Prinzips „Human using Language, Artefacts, Methodology, in which he is Trained“ (H-LAM/T) von Engelbart unterstützen.⁵⁵ Dabei geht es im Wesentlichen darum, dass Werkzeuge wie etwa Papier und Stift „allow humans in some sense to do what they already do in the way of thinking about problems – but do those things more efficiently – at least once they are trained in the use of these tools“.⁵⁶

Auch wenn in den beschriebenen Projekten die Interpretation der Texte kaum eine Rolle spielte, sind zwei ihrer Ergebnisse im Kontext der weiteren Betrachtung des gesamten Prozesses der hermeneutischen Textanalyse – und damit auch der Interpretation – relevant. Der beschriebene erweiterte hermeneutische Zirkel sowie der entwickelte Ablauf des Annotationsprozesses können nämlich als Ansätze für die Abbildung des gesamten hermeneutischen Textanalyseprozesses dienen. Auf dessen Basis können solche Tools entwickelt werden. Die Strategien zur Überprüfung nicht übereinstimmender Annotationen im entwickelten Modell zur kollaborativen Annotation können darüber hinaus – digital oder nicht-digital – als Zugang zu Interpretationsverfahren genutzt werden. Das zugrunde liegende Prinzip des „agreed disagreement“⁵⁷ könnte ein effektives Werkzeug zur Untersuchung von Interpretationsprozessen sein, da es zum einen auf klar strukturierten Analyseschritten basiert und zum anderen durch den kollaborativen Modus einen intersubjektiven Blick auf Interpretationen erzeugt. Außerdem erweitert es die hermeneutische Grundannahme „agreement is the basis of understanding“⁵⁸ unter Rückgriff auf das Konzept der textlichen Ambiguität bzw. Polyvalenz im Sinne der literaturwissenschaftlichen Paradigmen der Interpretationspluralität.

4.2.4 Vorverständnis

Im *Reallexikon* wird das ‚Vorverständnis‘ als Element diskutiert, das die gesamte hermeneutische Textanalyse beeinflusst „durch die Ansetzung eines

⁵⁵ Doug Engelbart Doug, *Augmenting Human Intellect. A Conceptual Framework* (1962), 11, zitiert nach Bradley (Anm. 53), 265.

⁵⁶ Ebd.

⁵⁷ Gius/Jacke (Anm. 31), 235.

⁵⁸ Dagfinn Føllesdal, „Hermeneutics and the Hypothetico-Deductive Method“, in: *Dialectica* 33/3–4 (1979), 319–336, hier: 336, <https://doi.org/10.1111/j.1746-8361.1979.tb00759.x>.

schon bestehenden oder sich spontan bildenden VORVERSTÄNDNISSES (d. h. einer heuristischen Antizipation) des Ganzen oder der im Text verhandelten Sache, das dann das Verstehen der Teile leitet und umgekehrt von diesem bestätigt oder revidiert wird“.⁵⁹ Weimar erläutert dabei nicht weiter, wie das Vorverständnis beschaffen ist. Da heuristische Verfahren immer dann eingesetzt werden, wenn das zur Verfügung stehende Wissen nicht ausreicht, um eine Frage zum Text zu beantworten, sind im Kontext der literaturwissenschaftlichen Textanalyse mit dem Vorverständnis als „heuristischer Antizipation“ vermutlich Annahmen gemeint, die beim Erschließen des Textverstehens eingesetzt werden können.

Im digitalen Zugang ist die Frage des Vorverständnisses ähnlich opak wie im nicht-digitalen. Allerdings wurde das Vorverständnis, das in den beiden Projekten in die hermeneutische Textanalyse mit einfluss, an manchen Stellen sichtbar. So wurde durch die Arbeit an und mit den Guidelines das terminologische Vorverständnis offensichtlich, das bei der Beschreibung des Textes eine Rolle spielt. Die Erstellung der Guidelines im Konfliktzählungsprojekt beinhaltete im ersten Schritt die schriftliche Fixierung des terminologischen Vorverständnisses. Mit diesem Vorverständnis wurde dann die Analyse versucht. In Zweifelsfällen oder Fällen, in denen sich ein Konzept nicht auf das vorhandene Textmaterial anwenden ließ, musste das Konzept so lange weiterentwickelt werden, bis es stabil anwendbar war. Damit kann die Begriffsarbeit auch als Entwicklung einer für die Beschreibung einsetzbaren Terminologie aus dem terminologischen Vorverständnis betrachtet werden.

Die kollaborative Arbeit mit den Guidelines im *heureCLÉA*-Projekt wiederum führte zur Explizierung weiterer impliziter Vorannahmen. Bei der Diskussion unterschiedlicher Annotationen wurde das (unterschiedliche) Vorverständnis der Annotatoren und Annatorinnen offengelegt. Dieses Vorverständnis betraf sowohl die Analysekonzepte als auch den Text selbst.

Eine dritte Form von ‚Vorverständnis‘ wurde schließlich durch die Entwicklung des Ablaufschemas für die literaturwissenschaftliche Annotation expliziert. Die systematische Beschreibung des Analyseprozesses ist als intersubjektiv gültige Offenlegung des Vorverständnisses von diesem Prozess intendiert. Auch wenn nicht eindeutig festgestellt werden kann, ob dies erreicht wurde, wurde zumindest das Vorverständnis der Projektbeteiligten expliziert und weiterentwickelt.

4.2.5 Weitere Aspekte

Die bisher beschriebenen Aspekte stellen einen Bezug zum literaturwissenschaftlichen Vorgehen her, das durch den digitalen Zugang verstärkt, aber nicht im Wesen verändert wurde. Es geht durchweg um Probleme der hermeneutischen Textanalyse, die durch den in Bezug auf Textmenge, Kategorienanzahl und Beteiligte umfangreicheren Zugang virulent wurden. Im Prinzip sind die resultierende Arbeit an Konzeptdefinitionen, die Methodenexplizierung und dergleichen natürlich nicht neu. Sie sind unabhängig vom digitalen Zugang möglich und notwendig.

⁵⁹Weimar (Anm. 39), 32.

Darüber hinaus gibt es aber einen Problemkomplex, der in der Literaturwissenschaft bisher nicht im Kontext der Verfahren der hermeneutischen Textanalyse diskutiert wurde. Hier geht es insbesondere um die bisher auch im Kontext der beschriebenen Projekte kaum thematisierte, bereits vor der eigentlichen Textanalyse liegende Frage nach der Qualität der Gegenstände, also nach den Primärtexten.

Die Qualität von Primärtexten wird bei einem großen Teil literaturwissenschaftlicher Analysen hauptsächlich anhand von Kriterien der Texttreue im weiteren Sinne (geeignete Textausgaben etc.) diskutiert. Die Zusammenstellung von verschiedenen Primärtexten zu einem Korpus hingegen wird in der Literaturwissenschaft zwar praktiziert, aber kaum problematisiert. Wenn die Forschungsfrage nicht zwingend mit der Textbasis zusammenhängt (z. B. bei einer Untersuchung aller Dramen Goethes in Bezug auf das Verhältnis zwischen männlichen und weiblichen Figuren), erfolgt die Korpuserstellung im nicht-digitalen Normalfall meist unter Verweis auf die Kanonizität der Texte. Die Exemplarität der Texte wird nicht weiter begründet, ihre Repräsentativität nur am Rande reflektiert. Auch in der Narratologie ist dies eine gängige Praxis. So begründet Genette (1998) in der Einführung zu seinem Standardwerk *Die Erzählung* die Analyse von Prousts *Auf der Suche nach der verlorenen Zeit* damit, dass er es analysiere und damit gleichzeitig seine Methode entwickle und anwende. Letzteres sei möglich, da die Besonderheiten von Prousts Text auch verallgemeinerbar seien.⁶⁰

In den vorgestellten Fällen ist dies anders: Bei der Untersuchung der Konflikt-narrative handelt es sich um ein literaturwissenschaftlich untypisches Korpus, da es erst generiert werden musste. Damit ist die Repräsentativität der Texte irrelevant, es können höchstens Kriterien für die Repräsentativität der erhobenen Interviews angelegt werden, was allerdings kein literaturwissenschaftliches, sondern ein sozialwissenschaftliches Problem ist.

Im *heureCLÉA*-Projekt liegt hingegen eher der Standardfall für eine literaturwissenschaftliche Textanalyse vor: Die Forschungsfrage bezieht sich auf bestehende Texte. Allerdings bedingt im Fall von *heureCLÉA* nicht das Textmaterial die Frage nach der automatisierten Analyse von Zeitphänomenen. Die Forschungsfrage ist vielmehr allgemeinerer, konzeptueller Natur – und damit vorerst unabhängig von konkreten Texten. Obwohl es insbesondere im digitalen Bereich grundsätzlich möglich ist, die Gesamtmenge bzw. Grundgesamtheit (*Population*) relevanter Texte in eine vollständige Datensammlung zu überführen, ist im Falle von *heureCLÉA* unklar, was die Texte ausmacht, die dieser Grundgesamtheit angehören.⁶¹ Aus pragmatischen Gründen wurden deshalb folgende Ansprüche an die Texte gestellt: Die Texte mussten frei verfügbar sein, einen Umfang von etwa 5–10 Seiten haben und nach 1800 entstanden sein, um

⁶⁰Das ist allerdings weniger eine Begründung der Zulässigkeit des Vorgehens als ein Indiz für die hinter dieser Behauptung steckende Annahme, ein allgemeingültiges Methodeninventar entworfen zu haben, das unabhängig vom konkreten literarischen Text sei.

⁶¹Für eine Übersicht zur Korpuserstellung im *Digital-Humanities*-Kontext vgl. Christof Schöch, „Aufbau von Datensammlungen“, in: Jannidis/Kohle/Rehbein (Anm. 10), 223–233, hier: 225 f.

die Sprachvarianz und semantischen Verschiebungen gering zu halten. Außerdem sollten sie im Sinne der Fragestellung möglichst viele zeitliche Informationen enthalten und eine einfache narrative Struktur aufweisen.

Diese Kriterien wurden in diversen Vorträgen und Publikationen genannt, sie haben aber bisher keine kritische Reaktion erzeugt. Dies zeugt von einer Leerstelle der literaturwissenschaftlichen Methodik in Bezug auf fachliche Routinen zur Zusammenstellung eines Textkorpus als Untersuchungsgegenstand. Diese Routinen sollten insbesondere jene Fälle beschreiben, in denen die Zusammenstellung der Texte nicht vollständig und eindeutig aus der Fragestellung abgeleitet werden kann. Dort stellt sich nämlich immer die Frage nach der Repräsentativität der ausgewählten Texte, und diese ist literaturwissenschaftlich bislang wenig beleuchtet. Dabei ist insbesondere die Vermeidung von Fehlschlüssen, die durch – unbemerkte – Korrelationen entstehen, eine große Herausforderung.⁶² So kann z. B. im Falle eines Korpus, anhand dessen die stilometrische Bestimmung des Autor-Genders vorgenommen werden soll, gleichzeitig eine weitgehend eindeutige Aufteilung der Genres der enthaltenen Texte entlang des Autor-Genders vorliegen. Bei der entwickelten, weitestgehend erfolgreichen Bestimmung dessen sind es dann aber nicht die *Gender*-Signale, sondern die *Genre*-Signale, die – versehentlich – gemessen werden.⁶³ Diesen Fehler zu entdecken, ist keine triviale Aufgabe.

Im Bereich der *Digital Humanities* gibt es durchaus Ansätze für die Erstellung und Bewertung von Korpora.⁶⁴ Die Beschreibung von unvollständigen Korpora in ihrer Funktion als Gegenstand literaturwissenschaftlicher Textanalyse steht allerdings – insbesondere für Fälle, die nicht als Referenzkorpora intendiert sind

⁶²Die in automatisierten Verfahren außerdem problematischen Korrelationen, die durch eine ungeeignete Operationalisierung entstehen, sind weder für das digitale *Close Reading* noch für die traditionelle Literaturwissenschaft ein (neues) Problem. Ihre Verhandlung in methodischen oder methodologischen Diskussionen ist Bestandteil des literaturwissenschaftlichen Forschungsparadigmas.

⁶³Laura Mandell, „Gender and Cultural Analytics. Finding or Making Stereotypes?“, in: Matthew K. Gold/Lauren F. Klein (Hg.), *Debates in Digital Humanities 2019*, Minneapolis 2019.

⁶⁴Vgl. z. B. aktuell den Beitrag von Ulrike Henny-Krahmer/Frederike u. a., „Criteria for Reviewing Digital Text Collections, version 1.0“, in: *IDE* (Februar 2017), <https://www.i-d-e.de/publikationen/weitereschriften/criteria-text-collections-version-1-0/> [letzter Aufruf 15.9.2017], der u. a. die Arbeiten von Winne (Martin Winne, *Developing Linguistic Corpora. A Guide to Good Practice*, Oxford 2005, <http://ota.ox.ac.uk/documents/creating/dlc/> [letzter Aufruf 15.9.2017]) und Schreibman u. a. (Susan Schreibman/Laura Mandell/Stephen Olsen, „Evaluating Digital Scholarship“, in: *Profession* [2011], 123–201) fortführt sowie die Evaluationskriterien für Editionen von Sahle u. a. mit aufnimmt (Dies., *Kriterien für die Besprechung digitaler Editionen, Version 1.1* [Juni 2014], <https://www.i-d-e.de/publikationen/weitereschriften/kriterien-version-1-1> [letzter Aufruf 15.9.2017]). Die Umsetzung der Kriterien erfolgte bereits in der sechsten Ausgabe der Rezensionsschrift *RIDE*: Ulrike Henny-Krahmer/Frederike Neuber, „Editorial. Reviewing Digital Text Collections“, in: *RIDE* 6 (2017), <https://doi.org/10.18716/ride.a.6.0>.

– noch aus.⁶⁵ Besonders für Ansätze, in denen der manuellen Annotation ein hoher Stellenwert zukommt und mit viel Aufwand betrieben wird, ist die Qualität des Korpus eine wesentliche Gelingensbedingung, die den vergleichsweise hohen Einsatz von Ressourcen erst rechtfertigt. Ein literaturwissenschaftlich nicht adäquates Korpus führt dazu, dass im Zweifelsfall mit viel Zeit und Mühe eine gute Forschungsfrage methodisch vorbildlich an einem dafür ungeeigneten Gegenstand untersucht wird. Damit wird nicht nur die Forschungsfrage nicht beantwortet bzw. nicht beantwortbar. Auch die Nachnutzbarkeit der Daten, die den Ressourceneinsatz rechtfertigt, ist aus literaturwissenschaftlicher Sicht nicht gegeben.

Ein weiterer Aspekt, der an die Frage der Korpuserstellung anschließt und mitbestimmt, inwiefern dieser Frage im Forschungsprozess größere Aufmerksamkeit gewidmet werden wird, ist die mit der Arbeit an Korpora verbundene wissenschaftliche Reputation. In der Literaturwissenschaft wird zwar die Erarbeitung von Editionen als literaturwissenschaftliche Tätigkeit honoriert – sie gelten als Publikationen und haben häufig die Bedeutung einer Monographie –, die Zusammenstellung und Aufbereitung eines Korpus wird allerdings bislang kaum als wissenschaftliche Leistung wahrgenommen, ebenso wenig wie die Anreicherung des Korpus mit Metadaten und Annotationen. Da außerdem sowohl die Korpuserstellung als auch die Auszeichnung der Korpustexte maßgeblich für die weitere Analyse der Texte sind, ist hier eine zweifache Entwicklung wünschenswert: Erstens braucht es die bereits geforderten Kriterien für die Korpuserstellung sowie Qualitätskriterien für die Textauszeichnung. Nur die Etablierung solcher Kriterien kann zweitens dazu führen, dass die mit der Korpuserstellung und -annotation verbundenen Tätigkeiten als wissenschaftliche Leistung anerkannt werden. Die Arbeit an und mit Korpora würde auf dieser Basis als wissenschaftliche Forschungs- und Publikationstätigkeit gewertet werden können. Vor allem kann sie infolgedessen im wissenschaftlichen Diskurs auf eine vergleichbare Weise adressiert werden wie fachlich etablierte Beitragsformen, so etwa Publikationen zu Textinterpretationen oder Methodenarbeiten. Im Kontext von Korpora und Korpusannotation besteht also sowohl in der digitalen als auch in der nicht-digitalen Literaturwissenschaft Bedarf an der Entwicklung von fachlich adäquaten Standards und Methoden.

⁶⁵Allerdings gibt es Überlegungen zum Korpusaufbau, die auch die nachfolgende literaturwissenschaftliche Analyse im Blick haben, u. a. bei: Alexander Geyken/Susanne Haaf/Bryan Jurish u. a., „Das Deutsche Textarchiv. Vom historischen Korpus zum aktiven Archiv“, in: Silke Schomburg/Claus Leggewie/Henning Lobin u. a. (Hg.), *Digitale Wissenschaft – Stand und Entwicklung digital vernetzter Forschung in Deutschland*. 20./21. September 2010, Köln. *Beiträge der Tagung*, Köln 2011, 157–161; Berenike Hermann/Gerhard Lauer, „Das ‚Was-bisher-geschah‘ von KOLIMO. Ein Update zum Korpus der literarischen Moderne“, in: *DHd 2017. Digitale Nachhaltigkeit. Konferenzabstracts* (2017), 107–111.

5 Fazit

Wie verhält es sich nun mit computergestütztem *Close Reading* und der hermeneutischen Textanalyse? Der Beitrag macht deutlich, dass es durchaus Abweichungen zwischen den digitalen und den nicht-digitalen Verfahren gibt. Je nachdem, ob diese Abweichungen als bereits in der literaturwissenschaftlichen Textanalyse angelegt oder als kategorial neue und damit potenziell widersprüchliche Elemente betrachtet werden, muss auch die Beurteilung des digitalen *Close Reading* als literaturwissenschaftliches Verfahren ausfallen. Computergestütztes *Close Reading* kann entsprechend als Intensivierung oder Vertiefung des hermeneutischen Zugangs oder eben als dessen Veränderung aufgefasst werden.

Die in diesem Beitrag erfolgte Ausrichtung der Darstellung der Effekte digitalen *Close Reading* an den Elementen literaturwissenschaftlicher Textanalyse soll die Integrierbarkeit des digitalen Zugangs demonstrieren und ist deshalb ein Plädoyer für die Betrachtung von digitalem *Close Reading* als literaturwissenschaftliches Forschungsparadigma. Das ist aber kein zwingender Schluss, denn man darf nicht vergessen: Die diskutierten Punkte sind ein Versuch, aus der Erfahrung einiger weniger, methodisch relativ homogener Forschungsprojekte heraus über die Konsequenzen nachzudenken, die die manuelle digitale Annotation für den hermeneutischen Textanalyseprozess und die damit verbundenen Praktiken der Korpuserstellung und Textannotation hat. Entsprechend basieren die gemachten Beobachtungen auf exemplarischen Anwendungsfällen.

Allerdings wurden Konsequenzen des computergestützten Ansatzes diskutiert, die allgemeiner gültig sein können. Betrachtet man nämlich jene Aspekte des digitalen *Close-Reading*-Prozesses, die für oben beschriebene Erweiterungen von literaturwissenschaftlichen Verfahren ursächlich sind, so scheinen diese unabhängig von den verfolgten Fragestellungen oder der angewandten Methode zu sein.

Insgesamt sind es drei Besonderheiten des digitalen Zugangs, die einander bedingen und – einzeln oder kombiniert – den literaturwissenschaftlichen Reflexionsprozess in den beiden Projekten angestoßen haben:

Annotation als distinktes Verfahren Annotationen im digitalen Kontext sind technisch bedingt genauer als nicht-digitale Annotationen. Die Annotation eines Textes erfordert die präzise Bestimmung der zu annotierenden Textstelle – selbst wenn es sich um einen *Textbereich* handelt, muss dieser auf das Zeichen genau definiert sein, um vom Computer gespeichert und weiterverarbeitet werden zu können. Ebenso ist der Umgang mit den Kategorien anders als im analogen Bereich: Die Vergabe von Tags erfordert eine vorherige Benennung dieser Tags und fördert eine anschließende Überprüfung der Stringenz und Plausibilität ihrer Vergabe. Analysekategorien müssen deshalb zwar nicht zwingend systematisch verwendet werden, aber ihre unsystematische Verwendung wird schneller offensichtlich.

Textanalyse als definierter Prozess Die beiden Effekte digitaler Annotation – genauere Text- und Konzeptbestimmung – führen wiederum dazu, dass der

Analyseprozess selbst in den Fokus der Analysierenden tritt. Auch hier gilt wieder, dass dies nicht gezwungenermaßen zu einer Explizierung und Systematisierung des Vorgehens führt. Allerdings ist es wahrscheinlicher, weil die bestehenden Defizite im Analyseprozess – implizite, widersprüchliche oder ungenaue Schritte – augenscheinlich werden und es naheliegend ist, diese zu beheben oder zumindest dadurch zur Prozessbeschreibung beizutragen, dass Gründe für die Nichtbehebung der Defizite angegeben werden.

Kollaboration als intersubjektiver Zugang Schließlich befördert die Tatsache, dass Annotationstools meistens kollaborative Funktionen haben, den ohnehin schon vorwiegend teamorientierten Arbeitsmodus in digitalen Projekten. Kollaboration stellt damit die dritte Besonderheit des digitalen Zugangs dar. Bereits das Speichern der Annotationen ermöglicht intra- und intersubjektiv eine intensivere Auseinandersetzung mit den in den Annotationen festgehaltenen Analysen. Kommt der kollaborative Modus hinzu, wird das intersubjektive Paradigma zum *default*-Fall in der Analyse und fördert nicht nur den Erkenntnisgewinn über Konzepte, Analyseverfahren und schließlich Gegenstände, sondern auch die Auseinandersetzung mit dem in der hermeneutischen Textanalyse stattfindenden Erkenntnisgewinn selbst.

Alle drei Besonderheiten digitaler Annotation sind, wie ausgeführt, unabhängig vom narratologischen Zugang. Deshalb kann man davon ausgehen, dass sie auf andere textanalytische Verfahren einen ähnlichen Einfluss haben. Unabhängig von der Frage der Passung zum literaturwissenschaftlichen Methodenparadigma scheinen daher digitale *Close-Reading*-Techniken generell als Ansatz geeignet, den Prozess der hermeneutischen Textanalyse aus der Praxis heraus zu erforschen und die Verfahren einer bestimmten Methode oder Interpretationstheorie offenzulegen, zu überprüfen und zu verbessern. Die Frage nach der Art des Zugangs – manuell oder automatisiert – ist dabei sekundär.

Ergänzung 2021

Seit dem Symposium „Digitale Literaturwissenschaft“ im Herbst 2017 gab es in den Feldern, die in meinem Beitrag adressiert werden, einige Entwicklungen. Auf der grundsätzlichen Ebene hat die – von mir am Rande diskutierte – Frage der Replizierbarkeit literaturwissenschaftlicher Forschung⁶⁶ in Form des als Replizierbarkeitsstudie verpackten Beitrags von Nan Z. Da⁶⁷ einen teilweise erbitterten Schlagabtausch über die vermeintlichen Schwachstellen der *Computational Literary Studies* ausgelöst. Die Frage nach dem Verhältnis zwischen dem digitalen

⁶⁶Für eine Übersicht zu Replizierbarkeitsstudien vgl. Christof Schöch/Karin van Dalen-Oskam/Maria Antoniak u. a., „Replication and Computational Literary Studies“, in: *Digital Humanities Conference 2020* (2020). <https://doi.org/10.5281/zenodo.3893428>.

⁶⁷Nan Z. Da, „The Computational Case against Computational Literary Studies“, in: *Critical Inquiry* 45/3 (2019), 601–39. <https://doi.org/10.1086/702594>.

und nicht-digitalen Feld der Literaturwissenschaft ist also weiterhin virulent und der Ausgang der Debatte offen.

Innerhalb des digitalen Feldes hingegen gibt es im Bereich der digitalen Hermeneutik hauptsächlich zwei positive Entwicklungen, die mit den Arbeiten in diesen Feldern einhergehen. Erstens hat insbesondere die Perspektive auf manuelle Annotation als literaturwissenschaftliche Textanalyse deutlich an Bedeutung gewonnen. Die Rolle manueller Annotationen in der hermeneutisch fundierten Datenerzeugung wird etwa in mehreren Beiträgen in den Sammelbänden von Nantke & Schlupkoth (2020) und – z. T. mit einer auf andere Geistes- und Sozialwissenschaften erweiterten Perspektive – Reiter, Pichler & Kuhn (2020) diskutiert. Außerdem scheint sich im literaturwissenschaftlichen Kontext eine Wahrnehmung von Annotationsrichtlinien als wissenschaftliche Ergebnisse zu entwickeln: Diese werden nicht nur häufiger publiziert⁶⁸, sondern die Fachzeitschrift *Cultural Analytics*⁶⁹ hat mittlerweile auch einen Sonderband angenommen und publiziert, der Annotationsrichtlinien enthält, die im Rahmen eines *Shared Task* zur Richtlinienerstellung für Erzählebenen entwickelt wurden.⁷⁰ Damit gibt es nun die erste, durch *Peer Reviewing* qualitätsgesicherte Publikation von Annotationsrichtlinien.

Die zweite Entwicklung betrifft die Frage der literaturwissenschaftlichen Korpuskompilierung, welche in den vergangenen Jahren ebenfalls prominenter wurde. Neben den Ansätzen zur Korpuserstellung, die Algee-Hewitt & McGurl (2015) unter Rückgriff auf kanonische und nicht- bzw. antikanonische Auswahllisten vorgestellt haben, gibt es inzwischen Überlegungen zur Kompilierung großer und sehr großer Korpora von Herrmann & Lauer (2018)⁷¹, die mit Blick auf die Operationalisierung literaturwissenschaftlicher Fragestellungen umgesetzt wurden, sowie größere Initiativen zur Erstellung literarischer Korpora. Besonders hervorzuheben sind hier die *European Literary Text Collection* (ELTeC) (Odebrecht, Burnard & Schöch 2021) sowie das in *DraCor* umgesetzte Konzept der *Programmable Corpora* (Fischer u. a. 2019), die beide literarische Korpora in mehreren Sprachen zur Verfügung stellen. Ähnlich wie das Erstellen von Annotationsrichtlinien hat dadurch auch das Kompilieren von Korpora eine Aufwertung als wissenschaftliche Leistung erfahren. Dies ist nicht

⁶⁸Vgl. dazu etwa die in der Zenodo Community „SPP CLS“ des aktuell laufenden Schwerpunktprogramm Computational Literary Studies publizierten Richtlinien: https://zenodo.org/communities/spp_cls/ (letzter Aufruf 27.08.2021).

⁶⁹Vgl. <https://culturalanalytics.org/> (letzter Aufruf 27.08.2021).

⁷⁰Vgl. Evelyn Gius/Nils Reiter/Marcus Willand, *A Shared Task for the Digital Humanities. Special Issue (Journal of Cultural Analytics)*. Bd. 2, in: *Journal of Cultural Analytics* 1 (2019), <https://culturalanalytics.org/issue/2254-vol-2-issue-1-2019>.

⁷¹Darauf aufbauend vgl. auch Evelyn Gius/Katharina Krüger/Carla Sökefeld, „Korpuserstellung als literaturwissenschaftliche Aufgabe“, in: *DHd 2019. Digital Humanities. Multimedial & multi-modal. Konferenzabstracts* (2019), 164–66.

zuletzt ersichtlich an der Nutzung der *CRedit* (Contributor Roles Taxonomy)⁷² im Bereich der *Digital Humanities*, in der *Data Curation* als Rolle definiert ist, oder in der erfolgreichen Etablierung von Datenjournalen wie dem *Digital Humanities*-spezifischen *Journal of Open Humanities Data* (JOHD)⁷³.

Literatur

- Sämtliche digitale Referenzen wurden letztmalig am 15.9.2017 bzw. am 26.10.2021 eingesehen.
- Algee-Hewitt, Mark/McGurl, Mark, "Between Canon and Corpus. Six Perspectives on 20th-Century Novels", in: *Pamphlets of the Stanford Literary Lab* 8 (2015), <http://litlab.stanford.edu/LiteraryLabPamphlet8.pdf>.
- Bögel, Thomas/Gertz, Michael/Gius, Evelyn u.a., „Collaborative Text Annotation Meets Machine Learning. heureCLÉA, a Digital Heuristic of Narrative“, in: *DHCommons Journal*, 1 (Juli 2015), <https://doi.org/10.5281/zenodo.324059>.
- Bögel, Thomas/Gertz, Michael/Gius, Evelyn u.a., „Gleiche Textdaten, unterschiedliche Erkenntnisziele? Zum Potential vermeintlich widersprüchlicher Zugänge zu Textanalyse“, in: *DHd2015 Von Daten zu Erkenntnissen. Book of Abstracts*, Graz 2015, 151–158.
- Bögel, Thomas/Gius, Evelyn/Jacke, Janina u.a., „Beyond Pragmatics. Disciplinary Profits of Interdisciplinary Approaches“, in: *DH2015 Global Digital Humanities. Conference Abstracts*, Sydney 2015.
- Bögel, Thomas/Gius, Evelyn/Jacke, Janina u.a., „From Order to Order Switch. Mediating between Complexity and Reproducibility in the Context of Automated Literary Annotation“, in: *Digital Humanities 2016 Conference Abstracts*, Krakow 2016, 379–382.
- Bögel, Thomas/Strötgen, Jannik/Gertz, Michael, „Computational Narratology. Extracting Tense Clusters from Narrative Texts“, in: *LREC Conference Proceedings*, 15. November 2014, http://www.lrec-conf.org/proceedings/lrec2014/pdf/280_Paper.pdf.
- Borek, Luise/Schöch, Christof/Thoden, Klaus, „Digitale Forschungsaktivitäten multilingual. TaDiRAH für die deutschsprachige DH-Community“, in: *DHd 2016 Modellierung, Vernetzung, Visualisierung. Die Digital Humanities als fächerübergreifendes Forschungsparadigma. Konferenzabstracts*, Leipzig 2016, 376–78.
- Bradley, John, „Thinking about Interpretation. Pliny and Scholarship in the Humanities“, in: *Literary and Linguistic Computing* 23/3 (2008), 263–279, DOI <https://doi.org/10.1093/lilc/fqn021>.
- Burnard, Lou, „On the Hermeneutic Implications of Text Encoding“, in: Domenico Fiormonte/ Jonathan Usher (Hg.), *New Media in the Humanities: Research and Applications. Proceedings of the First Seminar „Computers, Literature, and Philology“*, Edinburgh, 7–9 September 1998, Oxford 2001, 31–38.
- Da, Nan Z., „The Computational Case against Computational Literary Studies“, in: *Critical Inquiry* 45/3 (2019), 601–39. <https://doi.org/10.1086/702594>.
- Englebart, Doug, *Augmenting Human Intellect. A Conceptual Framework*, Summary Report, Stanford Research Institute, Menlo Park 1962.
- Fischer, Frank/Börner, Ingo/Göbel, Mathias u.a., „Programmable Corpora. Introducing DraCor, an Infrastructure for the Research on European Drama“, in: *Proceedings of DH2019* (2019). DOI <https://doi.org/10.5281/ZENODO.4284002>.
- Føllesdal, Dagfinn, „Hermeneutics and the Hypothetico-Deductive Method“, in: *Dialectica* 33/3-4 (1979), 319-336, DOI <https://doi.org/10.1111/j.1746-8361.1979.tb00759.x>.

⁷²Vgl. <https://casrai.org/credit/> (letzter Aufruf 27.08.2021).

⁷³Vgl. <https://openhumanitiesdata.metajnl.com/> (letzter Aufruf 27.08.2021).

- Genette, Gérard, *Die Erzählung*, München 1998.
- Geyken, Alexander/Haaf, Susanne/Jurish, Bryan u.a., „Das Deutsche Textarchiv. Vom historischen Korpus zum aktiven Archiv“, in: Silke Schomburg/Claus Leggewie/Henning Lobin u.a. (Hg.), *Digitale Wissenschaft – Stand und Entwicklung digital vernetzter Forschung in Deutschland. 20./21. September 2010, Köln. Beiträge der Tagung*, Köln 2011, 157–161.
- Gius, Evelyn, *Erzählen über Konflikte. Ein Beitrag zur digitalen Narratologie*, Berlin 2015.
- Gius, Evelyn, „Narration and Escalation. An Empirical Study of Conflict Narratives“, in: *Diegesis* 5/1 (2016), 4–25.
- Gius, Evelyn, *Korpus „Erzählen über Konflikte“* (August 2013), DOI <https://doi.org/10.5281/zenodo.894732>.
- Gius, Evelyn, *Narratologische Annotationen des Korpus’ „Erzählen über Konflikte“* (August 2013), DOI <https://doi.org/10.5281/zenodo.894720>.
- Gius, Evelyn/Jacke, Janina, „Informatik und Hermeneutik. Zum Mehrwert interdisziplinärer Textanalyse“, in: *Sonderband der Zeitschrift für digitale Geisteswissenschaften* 1 (2015), DOI https://doi.org/10.17175/sb001_006.
- Gius, Evelyn/Jacke, Janina, „Kollaboratives Annotieren literarischer Texte“, in: *DHd 2016. Modellierung, Vernetzung, Visualisierung. die Digital Humanities als fächerübergreifendes Forschungsparadigma. Konferenzabstracts*, Leipzig 2016, 240–243.
- Gius, Evelyn/Jacke, Janina, „The Hermeneutic Profit of Annotation. On preventing and fostering disagreement in literary text analysis“, in: *International Journal of Humanities and Arts Computing* 11/2 (2017), 233–254, DOI <https://doi.org/10.3366/ijhac.2017.0194>.
- Gius, Evelyn/Jacke, Janina, *Zur Annotation narratologischer Kategorien der Zeit. Guidelines zur Nutzung des CATMA-Tagsets*, Hamburg 2015, <http://heureclea.de/wp-content/uploads/2016/11/guidelinesV2.pdf>.
- Gius, Evelyn/Jacke, Janina/Meister, Jan Christoph u.a., *Heurecléa Source Documents: 1.0.* (2017), DOI <https://doi.org/10.5281/zenodo.274962>.
- Gius, Evelyn/Jacke, Janina/Meister, Jan Christoph u.a., *Heurecléa Time Annotations Compared Public: V1.1.* (22. Februar 2017), DOI <https://doi.org/10.5281/zenodo.595484>.
- Gius, Evelyn/Jacke, Janina/Meister, Jan Christoph, *Heurecléa Time Annotations Uncompared Public: V1.1.* (22. Februar 2017), DOI <https://doi.org/10.5281/zenodo.593119>.
- Gius, Evelyn/Krüger, Katharina/Sökefeld, Carla, „Korpuserstellung als literaturwissenschaftliche Aufgabe“, in: *DHd 2019. Digital Humanities. multimedial & multimodal. Konferenzabstracts* (2019), 164–66.
- Gius, Evelyn/Reiter, Nils/Willand, Marcus, A Shared Task for the Digital Humanities. Special Issue (Journal of Cultural Analytics). Bd. 2, in: *Journal of Cultural Analytics* 1 (2019), <https://culturalanalytics.org/issue/2254-vol-2-issue-1-2019>.
- Henny-Krahmer, Ulrike/Neuber, Frederike u.a., „Criteria for Reviewing Digital Text Collections, version 1.0“, in: *IDE* (Februar 2017), <https://www.i-d-e.de/publikationen/weitereschriften/criteria-text-collections-version-1-0/>.
- Henny-Krahmer, Ulrike/Neuber, Frederike, „Editorial. Reviewing Digital Text Collections“, in: *RIDE* 6 (2017), DOI <https://doi.org/10.18716/ride.a.6.0>.
- Hermann, Berenike/Lauer, Gerhard, „Das ‚Was-bisher-geschah‘ von KOLIMO. Ein Update zum Korpus der literarischen Moderne“, in: *DHd 2017. Digitale Nachhaltigkeit. Konferenzabstracts* (2017), 107–111.
- Herrmann, J. Berenike/Lauer, Gerhard, „Korpusliteraturwissenschaft. Zur Konzeption und Praxis am Beispiel eines Korpus zur literarischen Moderne“, in: *Osnabrücker Beiträge zur Sprachtheorie* 92 (2018), 127–156.
- Hühn, Peter, „Event and Eventfulness“, in: Peter Hühn/John Pier/Wolf Schmid u.a. (Hg.), *the living handbook of narratology* (13. September 2013), <http://www.lhn.uni-hamburg.de/article/event-and-eventfulness>.
- Hühn, Peter/Pier, John/Schmid, Wolf (Hg.), *the living handbook of narratology*, Hamburg, <http://www.lhn.uni-hamburg.de/>.

- Jacke, Janina, „Is There a Context-Free Way of Understanding Texts? The Case of Structuralist Narratology“, in: *Journal of Literary Theory* 8/1 (2014), 118–139, DOI <https://doi.org/10.1515/jlt-2014-0005>.
- Jacke, Janina/Meister, Jan Christoph, „Pushing Back the Boundary of Interpretation. Concept, Practice and Relevance of a Digital Heuristic“, in: *Digital Humanities 2014 – Book of Abstracts*, Lausanne 2014, 264–266.
- Jannidis, Fotis/Kohle, Hubertus/Rehbein, Malte (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart 2017.
- Kindt, Tom/Müller, Hans-Harald, „Zum Verhältnis von Interpretation und Deskription. Ein Bestimmungsvorschlag und ein Beispiel“, in: Jan Borkowski/Stefan Descher/Felicita Ferder/Philipp D. Heine (Hg.), *Literatur interpretieren. Interdisziplinäre Beiträge zur Theorie und Praxis*, Münster 2015, 73–90.
- Lahn, Silke/Meister, Jan C., *Einführung in die Erzähltextanalyse*, Stuttgart/Weimar 2013.
- Lancashire, Ian, *Using TACT with electronic texts. A guide to text-analysis computing tools, version 2.1 for MS-DOS and PC DOS*, New York 1996.
- Lotman, Jurij, *The Structure of the Artistic Text*, Ann Arbor 1977.
- Mandell, Laura, „Gender and Cultural Analytics. Finding or Making Stereotypes?“, in: Matthew K. Gold/Lauren F. Klein (Hg.), *Debates in Digital Humanities 2019*, Minneapolis 2019, 3–26.
- McGann, Jerome, „Marking Texts of Many Dimensions“, in: Susan Schreibman/Raymond G. Siemens/John Unsworth (Hg.), *A Companion to Digital Humanities*, Malden, MA, 2004, 198–217.
- Meister, Jan C./Petris, Marco/Gius, Evelyn u.a., *CATMA* (Version v5.2, 18. Juli 2018), DOI <https://doi.org/10.5281/zenodo.1470119>.
- Moretti, Franco, „Conjectures on World Literature“, in: *New Left Review* 1 (2000), 54–68.
- Nantke, Julia/Schlupkothén, Frederik (Hg.), *Annotations in Scholarly Editions and Research: Functions, Differentiation, Systematization*, Berlin/Boston 2020. DOI: <https://doi.org/10.1515/9783110689112>.
- Odebrecht, Carolin/Burnard, Lou/Schöch, Christof, “European Literary Text Collection (ELTeC). April 2021 release with 14 collections of at least 50 novels“ (2021). DOI: <https://doi.org/10.5281/ZENODO.3462435>.
- Piez, Wendell, „Towards Hermeneutic Markup. An Architectural Outline“, in: *Digital Humanities 2010. Center for Computing in the Humanities, King’s College. Conference Abstracts* (2010), 202–205, <http://piez.org/wendell/papers/dh2010/>.
- Ramsay, Stephen, „Algorithmic Criticism“, in: Raymond G. Siemens/Susan Schreibman (Hg.), *A companion to digital literary studies*, Malden, MA, 2008, 477–491.
- Rapp, Andrea, „Manuelle und automatische Annotation“, in: Jannidis, Fotis/Kohle, Hubertus/Rehbein, Malte (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart 2017, 253–267.
- Reiter, Nils/Pichler, Axel/Kuhn, Jonas (Hg.), *Reflektierte algorithmische Textanalyse. Interdisziplinäre(s) Arbeiten in der CRETA-Werkstatt*, Berlin/Boston 2020. DOI: <https://doi.org/10.1515/9783110693973>.
- Renear, Allen H., „Text Encoding“, in: Susan Schreibman/Raymond G. Siemens/John Unsworth (Hg.), *A companion to digital humanities*, Malden, MA, 2004, 218–239.
- Sahle, Patrick, *Kriterien für die Besprechung digitaler Editionen, Version 1.1* (Juni 2014), <https://www.i-d-e.de/publikationen/weitereschriften/kriterien-version-1-1>.
- Schmid, Wolf, *Elemente der Narratologie*, Berlin/Boston 2008.
- Schöch, Christof, „Aufbau von Datensammlungen“, in: Fotis Jannidis/Hubertus Kohle/Malte Rehbein (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart 2017, 223–233.
- Schöch, Christof, „Wiederholende Forschung in den digitalen Geisteswissenschaften“, in: *DHD 2017. Digitale Nachhaltigkeit. Konferenzabstracts* (2017), 207–212, http://www.dhd2017.ch/wp-content/uploads/2017/02/Abstractband_ergaenz.pdf.

- Schöch, Christof/van Dalen-Oskam, Karin/Antoniak, Maria u.a., „Replication and Computational Literary Studies“, in: *Digital Humanities Conference 2020* (2020). DOI: <https://doi.org/10.5281/zenodo.3893428>.
- Schreibman, Susan/Mandell, Laura/Olsen, Stephen, „Evaluating Digital Scholarship“, in: *Profession* (2011), 123–201.
- Schreibman, Susan/Siemens, Raymond G. (Hg.), *A companion to digital literary studies*, Malden, MA, 2008.
- Schreibman, Susan/Siemens, Raymond G./Unsworth, John (Hg.), *A companion to digital humanities*, Malden, MA, 2004.
- Schreibman, Susan/Siemens, Raymond G./Unsworth, John (Hg.), *A New Companion to Digital Humanities*, Chichester, West Sussex, UK, 2016.
- Sperberg-McQueen, „Classification and its Structures“, in: Susan Schreibman/Raymond G. Siemens/John Unsworth (Hg.), *A companion to digital humanities*, Malden, MA, 2004, 161–176.
- Spree, Axel, „Interpretation“, in: Harald Fricke/Klaus Grubmüller/Jan-Dirk Müller u.a. (Hg.), *Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*, Berlin 2000, 168–172.
- Unsworth, John, *Scholarly Primitives. What methods do humanities researchers have in common, and how might our tools reflect this?*, London 2000, <http://people.brandeis.edu/~unsworth/Kings.5-00/primitives.html>.
- Vogt, Jochen, *Einladung zur Literaturwissenschaft. Mit einem Vertiefungsprogramm im Internet*, München/Paderborn 2008.
- Weimar, Klaus, „Hermeneutik“, in: Harald Fricke/Klaus Grubmüller/Jan-Dirk Müller u.a. (Hg.), *Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*, Berlin 2000, 25–29.
- Weimar, Klaus, „Hermeneutischer Zirkel“, in: Harald Fricke/Klaus Grubmüller/Jan-Dirk Müller u.a. (Hg.), *Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*, Berlin 2000, 31–33.
- Winko, Simone, „Methode, literaturwissenschaftliche“, in: Ansgar Nünning (Hg.), *Metzler-Lexikon Literatur- und Kulturtheorie*, Stuttgart 2008, 495–496.
- Winko, Simone, „Textanalyse“, in: Harald Fricke/Klaus Grubmüller/Jan-Dirk Müller (Hg.), *Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*, Berlin 2003, 597–601.
- Winne, Martin, *Developing Linguistic Corpora. A Guide to Good Practice*, Oxford 2005, <http://ota.ox.ac.uk/documents/creating/dlc/>.

Online-Ressourcen

- TaDiRAH – Taxonomy of Digital Research Activities in the Humanities/TaDiRAH, <http://tadirah.dariah.eu/vocab/sobre.php>.
- TaDiRAH – Taxonomy of Digital Research Activities in the Humanities/Annotating, <http://tadirah.dariah.eu/vocab/index.php?tema=22&/annotating>.
- CATMA, www.catma.de.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Historisch-kritische Ausgabe digital. Eine Reformulierung der neugermanistischen Edition

Rüdiger Nutt-Kofoth

1 Präskript

Der folgende Beitrag behandelt die neugermanistische *Digitale Edition* in ihren theoretischen Prämissen sowie ihren strukturellen und methodischen Gegebenheiten. Die technologischen Rahmen der *Digitalen Edition* werden auf dieser Basis mitbedacht; es stehen aber nicht technische Einzelheiten in Hinblick auf divergente Software-Grundlagen oder spezifische Tools im Zentrum des hier vorgelegten Diskussionsbeitrags. Er versteht sich daher als philologischer, nicht als informatikzentrierter Blick auf das in Rede stehende Objekt.¹ Dies erscheint deshalb angemessen, weil die philologisch-phänomenologische Perspektive helfen kann, die im Titel des Beitrags annoncierte Reformulierung der neugermanistischen Edition vor dem Hintergrund auch der Geschichte von Theorie und Praxis der weit über ein Jahrhundert alten Historisch-kritischen Ausgabe und ihrer historischen Errungenschaften und Standards anzugehen. Damit sollen die Gewinne, die die *Digitale Edition* erzielen kann,² nicht gegen die Printedition

¹ Siehe analog die Feststellung von Mittler/Rehbein: „Schon allein aus rein wissenschafts-ökonomischen Gründen ist es notwendig, zu verhindern, dass das Rad stets von Neuem erfunden wird[,] und stattdessen zu erreichen, dass die Technologie in den Hintergrund rückt und sich die (digitalen) Editoren auf die eigentlich editorische Arbeit konzentrieren können“, Elmar Mittler/Malte Rehbein, „Edition und Forschungsbibliothek. Chancen und Herausforderungen einer traditionsreichen Partnerschaft im digitalen Zeitalter“, in: *Bibliothek und Wissenschaft* 44 (2011), 9–21, hier: 19.

² Ein interdisziplinärer Überblick über abgeschlossene und laufende digitale Editionsprojekte bei Patrick Sahle, *A Catalog of Digital Scholarly Editions*, v 3.0, snapshot 2008 ff. (letzte Änderung 22.3.2017), <http://www.digitale-edition.de/> (letzter Aufruf 22.9.2017).

R. Nutt-Kofoth (✉)
Bergische Universität Wuppertal, Wuppertal, Deutschland
E-Mail: nuttkofo@uni-wuppertal.de

ausgespielt, sondern in ein produktives Verhältnis zu ihr gesetzt werden. Insofern versteht der Beitrag den Medienwandel, der sich durch das Digitale vollzogen hat und immer noch weiter vollzieht, für die Editorik nicht als Revolution, sondern als Evolution.³

2 Die *Digitale Edition* im literaturwissenschaftlichen Feld

Als 2007 das dreibändige *Handbuch Literaturwissenschaft* erschien, konnte es seinen zweiten Band *Methoden und Theorien* mit einem Großabschnitt *Textkritik und Textbearbeitung* eröffnen, in dem sich allein zwei Kapitel befanden: dasjenige zu *Editionsphilologie* und dasjenige zu *Computerphilologie*.⁴ Die nächsten Großabschnitte beschäftigten sich dann mit *Textanalyse und Textinterpretation*, *Textbewertung* sowie *Literaturgeschichtsschreibung*.⁵ In dieser Ordnung spiegelt sich nahezu eins zu eins die Sortierung der Germanistik in die drei „disziplinären Kerne[]“ ‚Editionsphilologie‘, ‚Literaturgeschichtsschreibung‘ und ‚Interpretation‘, die Peter Strohschneider und Friedrich Vollhardt 2002 in den *Mitteilungen des Deutschen Germanistenverbandes* ansetzten, wobei sie bekanntlich der Editionsphilologie einen „in legitimatorischer Hinsicht insgesamt vergleichsweise unangefochtenen Status“ zuschrieben und sie als ersten dieser drei Grundpfeiler positionierten.⁶ In der Ordnung des literaturwissenschaftlichen Feldes war durch die Struktur des *Handbuchs Literaturwissenschaft* nun die ‚Computerphilologie‘ mit der ‚Editionsphilologie‘ auf eine Ebene gestellt, und damit waren beide als Pfeiler der literaturwissenschaftlichen Grundlagenforschung situiert. Die Aufteilung der Wissensfelder, die die beiden Artikel im *Handbuch Literaturwissenschaft* bedienen, weist nun allerdings eine Überschneidung auf, die zwar den vorgegebenen Artikelrahmen geschuldet ist, die nun aber zugleich für die in Rede stehende Sachlage bezeichnend ist. Gemeint ist der Bereich der *Digitalen Edition*. Im *Editionsphilologie*-Artikel nur angerissen,⁷ nimmt er im

³Ob die Zuspitzung dieser Grundannahme in der Attribuierung bei Elena Pierazzo zutrifft, wird erst die Zukunft zeigen können; Pierazzo: „digital scholarly editing is a radical evolution (but not revolution) of print-based editing“ (Elena Pierazzo, *Digital Scholarly Editing. Theories, Models and Methods*, Farnham, Surrey/Burlington, VT, 2015, 208 [Hervorhebung von R.N.-K.]).

⁴Rüdiger Nutt-Kofoth, „Editionsphilologie“, in: Thomas Anz (Hg.), *Handbuch Literaturwissenschaft. Gegenstände – Konzepte – Institutionen*, 3 Bde., Bd. 2: *Methoden und Theorien*, Stuttgart/Weimar 2007, 1–27; Fotis Jannidis, „Computerphilologie“, in: ebd., 27–40.

⁵Anz (Anm. 4), 41–284: [Diverse Autoren], „Textanalyse und Textinterpretation“, 41–231; Simone Winko, „Textbewertung“, in: ebd., 233–266; Jörg Schönert, „Literaturgeschichtsschreibung“, in: ebd., 267–284.

⁶Peter Strohschneider/Friedrich Vollhardt, „Interpretation. Eine Einleitung in den Themen teil dieses Hefes“, in: *Mitteilungen des Deutschen Germanistenverbandes* 49/2 (2002): *Interpretation*, 98–102, hier: 100.

⁷Nutt-Kofoth (Anm. 4), 24.

Computerphilologie-Artikel nicht nur einen größeren Raum ein,⁸ sondern er wird auch zu einem Kerngeschäft des behandelten Teilfelds erklärt: „Die Erstellung digitaler Editionen ist eines der Hauptarbeitsgebiete für Computerphilologen.“⁹ Zehn Jahre später hat sich das Feld der ‚Computerphilologie‘ weiter konturiert und kann mit der nunmehrigen Namensgebung *Digital Humanities* eine Eigenständigkeit im Rahmen der Geisteswissenschaften beanspruchen, wie z. B. die gleichnamige rezente Einführung aus dem Jahr 2017 ausweist.¹⁰ Die *Digitale Edition* erhält in dieser Einführung innerhalb des Großabschnitts *Digitale Objekte* ein eigenes Kapitel. Zwar macht dieses Kapitel nur etwa ein Achtzehntel des Buches aus, doch gilt die *Digitale Edition* – ganz ähnlich wie zehn Jahre zuvor – als eines der „prominentesten Themen der Digital Humanities“.¹¹

Die Wertigkeit, die der *Digitalen Edition* in den *Digital Humanities* zukommt, hat offensichtlich mit zwei Aspekten zu tun: Zum einen kann das Arbeitsfeld der *Digital Humanities* seine Leistungsfähigkeit natürlich besonders gut an einem Bereich der literaturwissenschaftlichen Grundlagenforschung zeigen und so seine besondere Bedeutung für ein Kerngebiet der Geisteswissenschaften verdeutlichen. Damit erhalten die *Digital Humanities* eine hochqualitative Wertigkeit für die Literaturwissenschaft. Zum anderen ist die ‚Edition‘ allerdings ein Objekt, das aufgrund seiner Bestandteile und seiner Struktur den *Digital Humanities* vielfältige Erprobungs- und Entwicklungsmöglichkeiten der digitalen Technologie anbietet. Zu nennen sind aus den zentralen Anliegen der Edition nur die Bereiche Bild-Text-Bezüge, differente, auch zu parallelisierende Textansichten, nicht in gewöhnlicher, linearer Fließtextform, sondern durch spezifische Darstellungen sichtbar zu machende Variantenverhältnisse bzw. textgenetische Prozesse, interne und externe Verweise sowie Verknüpfungen mit Kontextmaterialien.

Es sind also die neuen technischen Möglichkeiten des Digitalen, die spezifische Ausprägungen der Edition im digitalen Medium erlauben. Verbunden mit diesen neuen technischen Möglichkeiten wurde in den Diskussionen der *Digital Humanists* die Ausstellung eines vollständigen qualitativen Wechsels, eines qualitativen Schubs, der die *Digitale Edition* von der Printedition in jeglichen Belangen unterscheidet. Sie findet sich gebündelt in Patrick Sahles Definition der *Digitalen Edition*: „Scholarly digital editions are scholarly editions that are guided by a digital paradigm in their theory, method and practice.“¹² Dieser Definition

⁸ Jannidis (Anm. 4), 36–38.

⁹ Ebd., 36.

¹⁰ Fotis Jannidis/Hubertus Kohle/Malte Rehbein (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart 2017. In den zwei Jahrzehnten zuvor wurde der Weg von der computergestützten Edition zu *Born-digital*-Editionen beschritten; siehe zu den Anfängen Roland Kamzelak (Hg.), *Computergestützte Text-Edition*, Tübingen 1999; Roland S. Kamzelak, „Edition und EDV. Neue Editionspraxis durch Hypertext-Editionen“, in: Rüdiger Nutt-Kofoth/Bodo Plachta/H.T.M. van Vliet u. a. (Hg.), *Text und Edition. Positionen und Perspektiven*, Berlin 2000, 65–80.

¹¹ Patrick Sahle, „Digitale Edition“, in: Jannidis/Kohle/Rehbein (Anm. 10), 234–249, hier: 237.

¹² Patrick Sahle, „What is a Scholarly Digital Edition?“, in: Matthew James Driscoll/Elena Pierazzo (Hg.), *Digital Scholarly Editing. Theories and Practices*, Cambridge, UK, 2016, 19–39, hier: 28.

liegt folgende Vorannahme zugrunde: „A digital edition cannot be given in print without significant loss of content and functionality.“¹³ In Anschlag gebracht ist also ein Mehrwert der *Digitalen Edition* mit Bezug auf ihren Inhalt und ihre Funktionalität, der in Verbindung mit einem spezifischen theoretisch, methodisch und praktisch begründeten Paradigma des Digitalen stehe. Doch gilt es gerade vor dem Hintergrund der Geschichte der neugermanistischen Printedition, die angeführten Elemente ‚Inhalt‘ und ‚Funktionalität‘ sowie die Bestandteile des annoncierten digitalen Paradigmas genauer in Hinblick auf ihre ursächlich digitale Spezifität zu betrachten.

3 Spezifischer Mehrwert der *Digitalen Edition*: Inhalt und Funktionalität

Ohne Zweifel darf für die Kategorien ‚Inhalt‘ und ‚Funktionalität‘ der *Digitalen Edition* ein spezifischer Mehrwert aufgrund ihrer medientechnologischen Leistungsfähigkeit gegenüber der Printedition zugesprochen werden. Dabei ist allerdings zwischen quantitativem und qualitativem Mehrwert zu differenzieren. Unter ‚quantitativem Mehrwert‘ sei ein Mehrwert verstanden, der sich auf einen größeren Umfang, eine zeitliche Beschleunigung bezieht, die die Nutzung eines Editionsteils angenehmer, praktischer, unaufwändiger und zeitsparender macht. Ein ‚qualitativer Mehrwert‘ meint einen solchen, der sich ursächlich aus einer bestimmten Darstellungs- oder Vermittlungsmöglichkeit der *Digitalen Edition* ergibt, die aufgrund der andersmedialen Bedingungen des Digitalen allein in der *Digitalen Edition* realisierbar und durch die Printedition grundsätzlich nicht einlösbar ist.

Für die Kategorie ‚Inhalt‘ trifft nun zunächst einmal der Quantitätsaspekt zu. *Digitale Editionen* können problemlos als Bucheditionen große Mengen an Texten oder Bildern aufnehmen. Das betrifft nicht nur Volltextwiedergaben etwa aller für die Edition relevanten Überlieferungsträger in Bild und Text, sondern auch Volltextwiedergaben von Kontextdokumenten, etwa zur Entstehung und (zeitgenössischen) Rezeption. Hinzu können umfangreiche Auszüge oder gar ebenfalls Volltextwiedergaben von Quellen, aber auch von literarischen Folien bzw. Referenztexten kommen. Zugleich können Erläuterungen ebenfalls umfassender angelegt werden, indem Verweistexte, zum Beispiel Wörterbuchartikel für sprachliche Erläuterungen oder Lexikonartikel für Sacherläuterungen, im Volltext – falls sinnvoll und rechtlich unbedenklich – in die Edition eingebettet werden. Dieser quantitative Inhaltsaspekt kann durch einen qualitativen Inhaltsaspekt ergänzt werden. Er betrifft vor allem die Integration solcher Dokumente in die Edition, die nicht zum Bereich der Text- oder der Bilddokumente gehören; letztere können ja prinzipiell auch durch die Printedition wiedergegeben werden. Andersmediale

¹³ Ebd., 27.

Dokumente mit Ton- oder Bewegtbildelementen können jedoch allein im digitalen Medium repräsentiert werden, weil das Printmedium kein Speicher- und Ausgabeformat für Audio- und Videodokumente bietet. Der qualitative Mehrwert der Kategorie ‚Inhalt‘ betrifft also solche Dokumente, die nicht dem statisch-visuellen Speicher- und Ausgabeformat der Printedition entsprechen. Insofern beruht dieser Mehrwert auf der Fähigkeit des Digitalen, andere nicht-statisch-visuelle Medienformate aufgrund der dem Digitalen inhärenten multimedialen Speicher- und Ausgabeoptionen zu inkorporieren. Dies wird nicht nur für den Erläuterungsteil einer Edition hilfreich sein, sondern kann die zentralen Objekte der Edition betreffen, nämlich genau dann, wenn das Werk oder Teile von ihm (auch) aus Ton- und/oder Filmdokumenten bestehen, es also nicht oder nicht nur schrifttextlich überliefert ist bzw. überhaupt nicht oder nicht nur schrifttextuell ist.¹⁴

Die Elemente der Kategorie ‚Funktionalität‘ lassen sich weniger deutlich als die des ‚Inhalts‘ in eine quantitative und eine qualitative Merkmalsgruppe differenzieren, weil sich hier die Zuordnung bei den je einzelnen Elementen¹⁵ als changierend erweist. So sind z. B. einfache *Suchen* nach Personen-, Orts- oder Werknamen bei einer Volltextsuche in einer *Digitalen Edition* nur quantitativ in Hinblick auf die Suchzeit gegenüber einer Printedition besser, wenn die Printedition ein entsprechendes Register von Personen-, Orts- oder Werknamen enthält. Die Printedition kann sogar in Hinblick auf die Suchergebnisse qualitätshaltiger sein, wenn sie indirekte Erwähnungen solcher Namen im Register ebenfalls erfasst und die *Digitale Edition* solche Stellen in einer einfachen Suche nicht auswirft, es sei denn, diese Stellen sind ebenfalls im Sinne einer registerartigen Erfassung entsprechend ausgezeichnet. Andererseits ermöglicht schon die einfache Volltextsuche in *Digitalen Editionen* oder – soweit die Suchparameter das erlauben – in Editionsteilen die beliebige Suche nach Wörtern, Begriffen, allemal jeglichen Zeichenketten, die der Benutzer bildet, einschließlich der Regulierung nach unterschiedlichen Schärfegraden der Suchen, was sämtlich eine qualitative Differenz zu den entsprechenden Möglichkeiten im Printmedium ausmacht. Herkömmliche Register und Indizes, die klassische Bestandteile der Printedition bilden, sind dann für die *Digitale Edition* nicht mehr nötig.

Die *Verweisoptionen* der *Digitalen Edition* unterscheiden sich auf quantitativer Ebene von denjenigen der Printedition dadurch, dass erstens editionsinterne Links die verknüpfte Stelle mit einem Klick ohne das umständlichere Blättern sichtbar machen, und zwar auch als Ansicht (in einem neuen Fenster o. ä.), die die Ursprungsansicht auf dem Bildschirm und damit vor den Augen des Nutzers

¹⁴Vgl. dazu Rüdiger Nutt-Kofoth, „Autorschaft, Werk, Medialität. Editionstheoretische Annäherungen an pluriautorschaftliche und plurimediale Werkkomplexe – mit einem germanistischen Blick auf das Phänomen Oper/Libretto“, in: Thomas Betzwieser/Norbert Dubowy/Andreas Münzmay u. a. (Hg.), *Perspektiven der Edition musikdramatischer Texte*, Berlin/Boston 2017, 25–38.

¹⁵Einige der im Folgenden behandelten Elemente werden z. B. bei Sahle (Anm. 12), 29 f. und Sahle (Anm. 11), 240, genannt – allerdings anders bewertet.

hält, was beim Verschlagen von Buchseiten im gleichen Band nicht möglich wäre. Zweitens erlauben editionsexterne Verweise bei Referenzen zunächst einen quantitativen Zuwachs, indem sie sich des knapperen Raums für Zitate in der Buchedition entledigen und die Referenzstelle als ausführlicheres Zitat in die Edition integrieren oder mit einem externen Link auf diese an einem anderen digitalen Ort verweisen können, sodass der Nutzer den Umgebungstext der referenzierten Textstelle bzw. den zugehörigen Volltext (auch Bild, Ton, Film) direkt vorliegen hat, ohne ihn nach der bibliographischen Angabe in der Printedition erst an einem anderen Ort (in einer Bibliothek o. ä.) gegenständlich aufsuchen zu müssen. Wirklich qualitativ wird der externe Link in der *Digitalen Edition* erst dann, wenn er auf ein sich im Laufe der Zeit veränderndes, etwa aktualisiertes, aber unter der gleichen Adresse aufrufbares digitales Dokument bzw. eine solche Webseite verweist. Hier kann er nämlich durch die Hyperlinkstrukturoption des Digitalen zu einem je anderen Ergebnis führen, während der Verweis einer Printedition auf ein anderes Printobjekt zum immer gleichen Resultat führt.

Die Möglichkeit der *sukzessiven Veröffentlichung* bzw. der *fakultativen Unabgeschlossenheit* ist nun ein vorherrschend qualitatives Element, allerdings nur insofern, als mit ihm nicht die Verführbarkeit einhergeht, ein digitales Editionsprojekt auf dauerhafte Vorläufigkeit zu stellen, also es breit anzulegen, obwohl die Mittel und Detailkonzeption nicht ausreichen, es zu beenden. Dauerhafte Projektruinen oder besser Projektfragmente sind für die Außenwahrnehmung nicht hilfreich. Die Nachbesserungsmöglichkeit bei neuen Forschungserkenntnissen ist allerdings ein nicht zu unterschlagender Gewinn. Doch können Printeditionen solche Nachbesserungen zu einzelnen Bänden einer Edition auch noch vornehmen, solange der letzte Band, der Korrekturen und Ergänzungen zu früheren Bänden aufnehmen kann, noch nicht erschienen ist. Erst ab diesem zeitlichen Publikationsmoment schlägt der quantitative Aspekt dieses Elements für die *Digitale Edition* in einen qualitativen Aspekt um. Dies kann auch schon für das inzwischen häufiger angewandte Verfahren gelten, von nahezu fertigen Editionsprojekten vorab eine Beta-Version zu publizieren, um zum einen den Stand des Projekts öffentlich zu machen und zum anderen Kritik für die endgültige Fassung berücksichtigen zu können.

Das vielgenannte Element der *Kollaboration* ist eines, das sich ganz eigentlich schon in allen Printeditionen jüngerer Datums findet, die sämtlich von Editions-teams aus mehreren oder gar von einer ganzen Reihe von Mitarbeitern erarbeitet werden. Neu für die *Digitale Edition* ist allerdings, dass sich Mitarbeiter nach dem *Crowdsourcing*-Prinzip über ein einfaches Einschreibsystem selbst anbieten und Teile der Edition übernehmen können. Ein besonders bekanntes Beispiel dafür ist das *Transcribe-Bentham*-Projekt.¹⁶ Ob dies tatsächlich ein echter qualitativer Wechsel ist oder ob es sich doch eher um eine quantitative Verbesserung handelt,

¹⁶Jeremy Bentham, *Transcribe Bentham*, hg. von Philip Schofield/Melissa Terras/Tom Couch, London 2010 ff., <http://blogs.ucl.ac.uk/transcribe-bentham/> (letzter Aufruf 22.9.2017).

wäre genauer abzuwägen, denn auch Printeditionen könnten Mitarbeitersuchen annoncieren und Editionsarbeiten an Interessenten vergeben.

Das allerdings wohl grundlegendste Element der Funktionalität einer *Digitalen Edition* ist ihre *Transmedialität*.¹⁷ In ihr liegt das wesentliche Spezifikum der *Digitalen Edition* überhaupt, das in das Schlagwort „Daten vor Medien“¹⁸ gefasst werden kann. Es ist auch das einzige Element in der Kategorie ‚Funktionalität‘, das allein und vorbehaltlos als qualitativer Mehrwert der *Digitalen Edition* bezeichnet werden kann. Mit ihr verbunden ist die Einschätzung, dass die Edition aus ihren Daten besteht, die wiederum eine Übertragung in verschiedenmediale Ausgabeformate (etwa Bildschirmdarstellung oder Print) ermöglichen. Solche Optionen kann die Printedition nicht aufweisen, weil die Bedingungen des Printmediums nur einen untrennbaren Zusammenfall von Daten und Ausgabeformat erlauben. Man wird die Annahme, dass die *Digitale Edition* bloß aus den Daten besteht,¹⁹ also als Datenbank figuriert, in dieser Absolutsetzung dennoch infrage stellen können. Die Benutzerschnittstelle einer *Digitalen Edition*, das *Interface*, ist nämlich keineswegs eine beliebig austauschbare Komponente, sondern sie allein ermöglicht erst die Wahrnehmung der Editions Inhalte, und es ist gerade die spezifische gestalterische Funktionalität der Editions Oberfläche, die die Editions Inhalte adäquat zugänglich macht und damit die Gesamtfunktionalität der *Digitalen Edition* nicht gering beeinflusst. Schon in der Printedition dienen ja etwa ein hochdifferenziertes Layout und im Speziellen die Typografie als wesentliche Transporteur von Informationen an den Editionsutzer.²⁰

4 Spezifischer Mehrwert? Das digitale Paradigma und die Frage von Theorie, Methodik und Praxis der Edition

Lässt sich der quantitative oder qualitative Mehrwert der *Digitalen Edition* in Hinblick auf die Kategorien ‚Inhalt‘ und ‚Funktionalität‘ nur aufgrund eines genauen Vergleichs mit den Gegebenheiten und Möglichkeiten der Printedition ermitteln, so ist die Frage nach dem qualitativen Wechsel der Editorik aufgrund des digitalen

¹⁷ Siehe grundsätzlich dazu Patrick Sahle, „Zwischen Mediengebundenheit und Transmedialisierung. Anmerkungen zum Verhältnis von Edition und Medien“, in: *editio* 24 (2010), 23–36.

¹⁸ Ebd., 29.

¹⁹ Siehe etwa Patrick Sahle, *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, 3 Bde.: Bd. 1: *Das typographische Erbe*, Bd. 2: *Befunde, Theorie und Methodik*, Bd. 3: *Textbegriffe und Recodierung*, Norderstedt 2013 hier: Bd. 2, 154 f.

²⁰ Vgl. zu Letzterem Rüdiger Nutt-Kofoth, „Typographie als Informationssystem. Zum Layout der neugermanistischen Edition“, in: Rainer Falk/Thomas Rahn (Hg.), *Typographie & Literatur*, Frankfurt a. M./Basel 2016, 349–368.

Paradigmas, das eben auf den Funktionalitäten des Digitalen beruht, erst vor dem Hintergrund der Geschichte der Editorik zu beantworten. Die in Anschlag gebrachten wissenschaftlichen Grundkategorien der Theorie, Methodik und Praxis sind also in dieser Hinsicht zu untersuchen.

Was das digitale Paradigma bewirken könnte, ist aus der Perspektive der *Digitalen Edition* im Grundlagenwerk von Patrick Sahle so beschrieben worden: „Letztlich erfordert der Übergang zu digitalen Editionsformen einen so allgemeinen Wandel der Grundannahmen und Zielsetzungen, dass die Gültigkeit der bisherigen Methodologie grundsätzlich in Frage gestellt scheint.“²¹ Edward Vanhoutte hat vor einem solchen Hintergrund schon vor etwa 20 Jahren insbesondere die germanistische Editorik als Innovationsverweigerer ausgemacht: „The theory and practice of the creation of an electronic edition is being countered by resistance from the side of the hard-copy focussed German school of *Editionswissenschaft*.“²² Vanhoutte hat dies mit der Annahme begründet, dass im Gegensatz zur Printedition in der *Digitalen Edition* „the dogma of the solid text as presented by the editor is undermined and replaced by the tolerant acceptance of a text as a contextualized possibility.“²³ Gerade in Hinblick auf den Textbegriff sieht auch Sahle die Innovationskraft der *Digitalen Edition*: „Damit haben technische Veränderungen letztlich einen grundlegenden Einfluss darauf, was wir eigentlich unter den zu edierenden Texten verstehen und wie wir mit ihnen umgehen. Technologien verändern unseren **Textbegriff**.“²⁴ An anderer Stelle hat er schlagwortartig formuliert: „Die Evolution der Techniken ist eine Evolution der Textbegriffe.“²⁵

Der historische Blick auf die *Theoriediskussion* zumindest der neugermanistischen Editorik und ihre Auswirkungen auf die Edition kann diese Annahmen allerdings nicht bestätigen. Schon 1924 hatte Reinhold Backmann, der Grillparzer-Editor, die Hierarchisierung von herausgehobener Textfassung im Volltext (sog. edierter Text) und Apparat mit den Varianten aller anderen Fassungen grundsätzlich infrage gestellt und dem Apparat „seinen selbständigen Wert gegenüber dem Textabdruck“, ja sogar „ein Übergewicht an Bedeutung über den letzteren“ zugesprochen.²⁶ In Fortführung Backmanns hatte Gunter Martens

²¹ Sahle (Anm. 19), Bd. 2, 86.

²² Edward Vanhoutte, „Where is the editor? Resistance in the creation of an electronic critical edition“, in: *Human IT 1* (1999), 197–214, <http://etjanst.hb.se/bhs/ith/1-99/ev.htm> (letzter Aufruf 22.9.2017), Kap. 5.

²³ Ebd.

²⁴ Sahle (Anm. 11), 240 [Hervorhebung im Original].

²⁵ Sahle (Anm. 19), Bd. 3, 391.

²⁶ Reinhold Backmann, „Die Gestaltung des Apparates in den kritischen Ausgaben neuerer deutscher Dichter. (Mit besonderer Berücksichtigung der großen *Grillparzer-Ausgabe* der Stadt Wien)“, in: *Euphorion* 25 (1924), 629–662, hier: 638. Der Beitrag ist teilweise wiederabgedruckt in Rüdiger Nutt-Kofoth (Hg.), *Dokumente zur Geschichte der neugermanistischen Edition*, Tübingen 2005, 115–137, hier 125.

1971 aus der Überlieferungslage für die Lyrik Georg Heyms die Theorie der Textdynamik entwickelt. Dabei verstand Martens „die über variante Stadien verlaufende Entwicklung eines Werkes [...] als eine wichtige textspezifische Aussage, ja sogar als spezifische Qualität von Text schlechthin.“²⁷ Wenn dann die „Textvarianz eine zentrale Rolle“ einnimmt („sie repräsentiert somit die Textdynamik“),²⁸ ist schon 1971 der vormalige Apparat zum „Kernstück“ der Edition erklärt; der edierte Text wird dagegen zum „Superadditum zur Variantendokumentation“ reduziert und nur noch als „notwendige Konzession an den Leser“ verstanden,²⁹ eine Konzession, die dann die *Heym-Ausgabe* aus dem Jahr 1993 auch nicht mehr als „notwendig[]“ verstand und auf die sie zugunsten der Repräsentation des Werktextes in rein zeilensynoptischen Darstellungen gar ganz verzichtete.³⁰

Spätestens in den 1980er Jahren war der Werkbegriff in diesen texttheoretischen Umschlag eingebunden. 1982 hatte Siegfried Scheibe definiert: „Der Text eines Werkes im editorischen Sinne besteht aus den Texten sämtlicher Textfassungen, die im Laufe des Entstehungsprozesses eines Werkes vom Autor oder in seinem Auftrag zu diesem Werk hergestellt wurden.“³¹ Dass der Werkbegriff später auch wieder mit der Öffentlichkeitsbestimmung verbunden oder auf die Veröffentlichung zugespitzt wurde,³² tut dieser theoretischen Ausfaltung des Textbegriffs keinen Abbruch. Wichtig für die Edition ist ja vor allem, dass nicht mehr der herkömmliche edierte Text, die Lesetextfassung einer Edition, als ausschließliche Repräsentation des Werkes begriffen war.

Zwei weitere Beispiele für texttheoretische Überlegungen, die im Bereich der neugermanistischen Printedition entwickelt wurden, seien noch kurz angeführt: Die eine ist Roland Reuß' Maxime ‚Ein Entwurf ist kein Text‘, die aus der Vorstellung hergeleitet ist, Text sei durch Linearität gekennzeichnet, weshalb ein Entwurf mit seinen nicht-linear platzierten paradigmatischen Elementen des ursprünglich Niedergeschriebenen (über, unter, neben die Zeile gesetzte

²⁷ Gunter Martens, „Textdynamik und Edition. Überlegungen zur Bedeutung und Darstellung variierender Textstufen“, in: Gunter Martens/Hans Zeller (Hg.), *Texte und Varianten. Probleme ihrer Edition und Interpretation*, München 1971, 165–201, hier: 167.

²⁸ Ebd., 170 f.

²⁹ Ebd., 171 f.

³⁰ Georg Heym, *Gedichte 1910–1912. Historisch-kritische Ausgabe aller Texte in genetischer Darstellung*, hg. von Günter Dammann/Gunter Martens/Karl L. Schneider, 2 Bde., Tübingen 1993.

³¹ Siegfried Scheibe, „Zum editorischen Problem des Textes“, in: *Zeitschrift für deutsche Philologie* 101 (1982), Sonderheft: Norbert Oellers/Hartmut Steinecke (Hg.), *Probleme neugermanistischer Edition*, 12–29, hier: 28.

³² Siegfried Scheibe, „Editorische Grundmodelle“, in: Siegfried Scheibe/Christel Laufer (Hg.), *Zu Werk und Text. Beiträge zur Textologie*, Berlin 1991, 23–48, hier: 25; Gunter Martens, „Das Werk als Grenze. Ein Versuch zur terminologischen Bestimmung eines editorischen Begriffs“, in: *editio* 18, 2004, 175–186, bes. 179 u. 184–186. Siehe auch die jüngere Rekapitulation bei Johnny Kondrup, „Text und Werk – zwei Begriffe auf dem Prüfstand“, in: *editio* 27 (2013), 1–14, zur Kritik des Martens'schen Werkbegriffs dort 9 f.

Änderungen) keinen Textstatus aufweise.³³ Die Konsequenz dieser Prämisse ist, dass die Edition nur noch Faksimiles und raummimetische Transkriptionen, aber keinen aus einem Entwurf konstituierten Text mehr enthalten soll.³⁴ Das andere Beispiel ist Herbert Krafts Verfahren der Fragmentedition. Es steht unter der Leitmaxime der „Räumlichkeit als Theorem der Fragmentedition“.³⁵ Da nach Kraft das „Fragmentarische [...] in seiner räumlichen Semantik“ erscheint und daher die „Anordnung eine strukturelle Wertigkeit besitzt“,³⁶ wird die Räumlichkeit zum Kern des Kraft'schen Verständnisses von der Fragmentedition. Sie ist in der Edition in Hinblick auf ihre strukturelle, nicht aber ihre zufällige Ausprägung in der Entwurfshandschrift wiederzugeben.³⁷ Anders als bei Reuß ist Krafts Textpräsentation in der Edition daher „von einer Nachbildung der Handschrift kategorial verschieden.“³⁸

So bleibt festzuhalten, dass die Infragestellung des edierten Textes als des Leittextes einer Edition zugunsten der Pluralität von Textdarstellungen keineswegs ein Charakteristikum der *Digitalen Edition* ist, sondern seit einem Jahrhundert in der neugermanistischen Editorik umfassend diskutiert und in den Printeditionen in einer reichen Anzahl an Facetten ausgeprägt ist.³⁹ Es wundert deshalb nicht, dass grundsätzliche Fragen an den Textbegriff gerade aus der neugermanistischen

³³ Roland Reuß, „Text, Entwurf, Werk“, in: *Text. Kritische Beiträge* 10 (2005): *Text · Werk*, 1–12, hier: 7.

³⁴ Realisiert z. B. in Franz Kafka, *Historisch-Kritische Ausgabe sämtlicher Handschriften, Drucke und Typoskripte*, hg. von Roland Reuß/Peter Staengle, Basel/Frankfurt a. M. 1995 ff. Vgl. zu dieser Fragestellung den Überblick bei Rüdiger Nutt-Kofoth, „Editorische Axiome“, in: *editio* 26 (2012), 59–71, hier: 65 f.

³⁵ So im Kapiteltitel in Herbert Kraft, *Editionsphilologie*, zweite, neubearbeitete und erw. Aufl. mit Beiträgen von Diana Schilling und Gert Vonhoff, Frankfurt a. M. u. a. 2001, 132; so auch in der ersten Aufl.: Herbert Kraft, *Editionsphilologie*, mit Beiträgen von Jürgen Gregolin u. a., Darmstadt 1990, 107.

³⁶ Kraft 2001 (Anm. 35), 135; auch in Kraft 1990 (Anm. 35), 111 f.

³⁷ Hierzu und zum literaturwissenschaftlichen Theoriezusammenhang von Krafts Konzept siehe Herbert Kraft, „Die Edition fragmentarischer Werke“, in: *LiLi. Zeitschrift für Literaturwissenschaft und Linguistik* 5/19–20 (1975), 142–146.

³⁸ Kraft 2001 (Anm. 35), 135; auch in Kraft 1990 (Anm. 35), 111. Zu Krafts und zu Reuß' Verständnissen der Fragmentedition und ihren Prämissen siehe ausführlicher Rüdiger Nutt-Kofoth, „Konzepte der Fragmentedition und ihre Probleme“, in: Matthias Berning u. a. (Hg.), *Fragment und Gesamtwerk. Relationsbestimmungen in Edition und Interpretation*, Kassel 2015, 13–27, hier: 16–20.

³⁹ Insofern trifft die folgende Kennzeichnung des Mehrwerts einer *Digitalen Edition* gerade nicht zu: „Der Abschied von der einen prämierten Textfassung ist ein unmittelbarer Effekt der neuen Technologien. Digitale Arbeits- und Publikationsformen fordern ein neues Nachdenken darüber, wie textliche Varianz zu verzeichnen und zu präsentieren ist. [...] Die neuen Medien führen zu einem neuen Textbegriff: der teleologische Text [...] wird jetzt von der Idee begleitet, dass ein Werk auch als ‚Summe der Fassungen‘ betrachtet werden kann“ (Sahle 2013 [Anm. 19], Bd. 2, 186 f.). Schon der für die Printedition entwickelte Begriff der ‚Summe der Fassungen‘ (siehe oben) ist ja geeignet, diese Annahme in Zweifel zu ziehen.

Editorik heraus im Zusammenhang mit den zeitgenössisch verfügbaren Printeditionen gestellt wurden.⁴⁰ Dies führte sogar dazu, dass ein Editionswissenschaftler den Grundlagenartikel *Text* für die zweite Auflage des *Reallexikons der deutschen Literaturgeschichte* verfasste.⁴¹

Wie aber sieht es mit der Feststellung eines grundsätzlichen ‚Shifts‘ auch für die *Methodik der Digitalen Edition* aus? Folgende Abfolge der Textordnung ist als Kennzeichen der *Digitalen Edition* genannt worden:

Eine umfassende Transkription nach Möglichkeit aller Textzeugen [...] ist die Basis für eine wissenschaftliche Edition und gehört unbedingt auch mit in die Publikation. Einen anderen Status hat dagegen heute eine kritische Fassung. Diese *kann* vom Editor erstellt werden, es haben aber auch Ausgaben als wissenschaftliche Editionen zu gelten, die auf eine solche Fassung verzichten und sich ganz auf die Erarbeitung der textlichen Grundlagen aus der Überlieferung und auf ihre kritische Erschließung konzentrieren. [...] Das Verhältnis der einzelnen Formen der Textwiedergabe hat sich im Vergleich zur Druckkultur damit umgekehrt, [...] bildet aber nun genau die Abfolge der editorischen Arbeitsschritte ab.⁴²

Schon der obige Abriss aus den Theorie-Diskussionen der neugermanistischen Editorik hat gezeigt, dass die in diesem Zitat genannten methodischen Verfahren der *Digitalen Edition* nicht ursächlich mit ihr zusammenhängen. Die printeditorische Diskussion um den Textbegriff etwa hat die Herstellung einer „kritische[n] Fassung“, also des sog. edierten Textes, je nach Textkonzept des Editors als ganz unterschiedlich gewichtig bis hin zu obsolet klassifiziert. Was die Methodik der editionsinternen Ordnung betrifft, so hat keineswegs die *Digitale Edition* eine Anordnung nach den editorischen Arbeitsschritten eingeführt. Die neugermanistische Printedition kennt diese schon seit 1975, als die Frankfurter *Hölderlin-Ausgabe* zu erscheinen begann.⁴³ Es ist diese Druckausgabe, die die Hierarchie vom edierten Text als Kopf der Edition und dem Apparat als nachgeordnetem Editionsteil aufgehoben hat, indem schon sie die Ordnung ‚Faksimile – Transkription – genetische Darstellung – konstituierter Text‘ anbot. Sie setzt dabei Hans Zellers Editionsformel der Trennung von Befund und Deutung in eine Abfolge um, die für die einzelnen Textdarstellungen eine sukzessive Abnahme an befundhaltigen und eine entsprechende Zunahme an deutungshaltigen Elementen aufweist. Die Befund-Deutung-Formel ist aber weit in der vor-

⁴⁰ Gunter Martens, „Was ist ein Text? Ansätze zur Bestimmung eines Leitbegriffs der Textphilologie“, in: *Poetica* 21, 1989, 1–25; der Beitrag erschien kurz darauf unwesentlich verändert als Gunter Martens, „Was ist – aus editorischer Sicht – ein Text? Überlegungen zur Bestimmung eines Zentralbegriffs der Editionsphilologie“, in: Scheibe/Laufer (Anm. 40), 135–156.

⁴¹ Gunter Martens, „Text“, in: Klaus Kanzog/Achim Masser (Hg.), *Reallexikon der deutschen Literaturgeschichte*, Bd. 4: *Sl–Z*, Berlin/New York ²1984, 403–417.

⁴² Sahle 2013 (Anm. 19), Bd. 2, 183.

⁴³ Friedrich Hölderlin, *Sämtliche Werke. ‚Frankfurter Ausgabe‘. Historisch-kritische Ausgabe*, hg. von D.E. Sattler, Frankfurt a. M. (seit 1985: Basel/Frankfurt a. M.) 1975–2008.

digitalen Zeit, nämlich 1971,⁴⁴ entwickelt worden und hat mit der Diskussion um die mediale Form der Edition nichts zu tun. Der Frankfurter *Hölderlin-Ausgabe* ist im Übrigen eine Anzahl an neugermanistischen Printeditionen gefolgt, die das Prinzip der gestaffelten mehrfachen Textwiedergabe ebenfalls umsetzte, wenn auch nicht immer – allerdings eben nicht aus medialen, sondern aus texttheoretischen Gründen (s. o.) – sämtliche Textsichten, etwa gelegentlich nicht die textgenetische und insbesondere nicht die textkonstituierende, zur Verfügung stellte.⁴⁵

Der Wandel der Editionsordnung gründet damit auf einer methodischen Einsicht, die der dokumentnahen wie der konstituierend-emendierenden Präsentation schon ihr gleichgewichtiges Recht für die Buchedition gegeben hat. Die damit verbundene Offenlegung der Editionsgrundlagen, die wiederum eine Offenheit für benutzerseitige Kritik impliziert, ist folgerichtig schon die Zielsetzung der Frankfurter *Hölderlin-Ausgabe* gewesen. Das Hölderlin-Zitat „Komm ins Offene, Freund!“ zum Abschluss des *Vorläufigen Editionsberichts* 1977 bildet ganz in diesem Sinne das Motto für die methodische Anlage der Edition, die als „offene editorische Darstellung [...] jedem Leser eine eigene Erkenntnis überläßt“.⁴⁶ Und selbst diese unautoritäre Editionshaltung gründet bloß auf die noch ein Jahrzehnt ältere Zielsetzung Hans Zellers: „Die Edition soll offen sein, sie hat nicht um jeden Preis eine Lösung, noch viel weniger eine runde Lösung zu geben, sondern die Möglichkeit zu Lösungen, indem sie das Material vermittelt.“⁴⁷ Auch diese Zielsetzung hat also mit der Entwicklung der *Digitalen Edition* und ihren Möglichkeiten zunächst einmal nichts zu tun, auch wenn sie gern – als Vorbereitung der Diskussion um die kollaborativen Optionen der *Digitalen Edition* – als ein Eigenwert des neuen Mediums eingebracht wird.⁴⁸

⁴⁴ Hans Zeller, „Befund und Deutung. Interpretation und Dokumentation als Ziel und Methode der Edition“, in: Martens/Zeller (Anm. 27), 45–89.

⁴⁵ Siehe neben den Ausgaben Heym (Anm. 30) und Kafka (Anm. 34) etwa die Ausgaben Heinrich v. Kleist, *Sämtliche Werke*, [Berliner, seit 1992:] *Brandenburger Ausgabe*, hg. von Roland Reuß/Peter Staengle, Basel/Frankfurt a. M. 1988–2010; Georg Trakl, *Sämtliche Werke und Briefwechsel*. *Innsbrucker Ausgabe*. *Historisch-kritische Ausgabe mit Faksimiles der handschriftlichen Texte* Trakls, hg. von Eberhard Sauermaun/Hermann Zwerschina, Basel/Frankfurt a. M. 1995–2014; Georg Büchner, *Sämtliche Werke und Schriften*. *Historisch-kritische Ausgabe mit Quellendokumentation und Kommentar* (Marburger Ausgabe), im Auftrag der Akademie der Wissenschaften und der Literatur, Mainz, hg. von Burghard Dedner, mitbegründet von Thomas Michael Mayer, Darmstadt 2000–2013.

⁴⁶ Wolfram Groddeck/D.E. Sattler, „Frankfurter Hölderlin-Ausgabe. Vorläufiger Editionsbericht“, in: *Le pauvre Holterling. Blätter zur Frankfurter Hölderlin-Ausgabe* 2 (1977), 5–19, hier: 19.

⁴⁷ Hans Zeller, „Edition und Interpretation. Antrittsvorlesung“, in: *zürcher student* 43/7 (Januar 1966), 15 u. 19, hier: 19; der Beitrag ist wiederabgedruckt in Nutt-Kofoth (Anm. 26), 279–288, hier: 288.

⁴⁸ Siehe etwa Sahle (Anm. 19), Bd. 2, 266: „Im Idealfall ist die digitale Edition vollständig transparent und jede editorische Entscheidung offengelegt. Der Nutzer kann alles nachvollziehen und dadurch ggf. auch zu anderen Schlüssen und editorischen Vorschlägen kommen. [...] Die Offenheit und Materialfülle der Edition kann den Nutzer zu einem anderen Verhalten inspirieren. Sie kann ihn vom passiven Leser zum aktiven Co-Editor machen“.

Von den drei in Rede stehenden wissenschaftlichen Grundkategorien Theorie, Methodik und Praxis kann am ehesten noch die Editionspraxis als Bestandteil eines digitalen Paradigmas gelten, das die *Digitale Edition* gegenüber der Printedition zu einem grundsätzlich anderen Editionstyp macht. Doch ist daran zu erinnern, dass ja nicht nur die methodische Frage der „**mehrfache[n] Textwiedergaben** auf der **Skala zwischen Quellennähe und Benutzernähe**“⁴⁹ von der Printedition schon erprobt wurde, sondern die mehrfache Textwiedergabe überhaupt. Das betrifft dann etwa diejenige von mehreren Fassungen als Volltext. Allein die Geschichte der Goethe-Philologie zeigt, wie dies schon in der *Weimarer Ausgabe* 1887–1919 realisiert wurde, nämlich als Abdruck in verschiedenen Bänden.⁵⁰ Der Paralleldruck von Fassungen fand ebenfalls schon immer Anwendung in kritischen Printausgaben, etwa in der *Goethe-Akademie-Ausgabe* der 1950/60er Jahre beim *Werther*⁵¹ oder bei *Faust I*. Bei letzterem Werk wurden drei Fassungen parallel produziert.⁵² Auch wurden noch mehr Fassungen schon vor weit mehr als einem Jahrhundert im Buch parallelisiert, was durchaus nicht zur Unübersichtlichkeit führte. Ein Beispiel ist die Edition von Goethes *Iphigenie auf Tauris* von 1883, in der vier Fassungen parallel erscheinen.⁵³

Neben der Frage der Textangebote kann ebenso die Aufgabe der Vernetzung zur Editionspraxis gezählt werden. Die Printedition hat auch in diesem Fall schon Lösungen für komplexe Fälle angeboten. So hat die Frankfurter *Hölderlin-Ausgabe* mit einer ausgefeilten Zählung von Manuskriptzeilen in der Marginalspalte der den Zeilenfall der faksimilierten Handschrift wiederpiegelnden Transkription die genetische Darstellung mit der Transkription verknüpft, sodass der Benutzer jede Einzelstelle zwischen den verschiedenen Textdarstellungen referenziert erhält und somit problemlos verbinden kann. Die Innsbrucker *Trakl-Ausgabe* zum Beispiel hat dieses Verfahren dann übernommen. Die Marburger *Büchner-Ausgabe* wiederum hat es sich u. a. zur Aufgabe gemacht, die Bezüge zwischen Autortext und Quellen detailliert nachzuweisen. Das tut sie nicht nur durch umfangreiche Abdrucke der Quellentexte, sondern auch sie arbeitet mit einer minutiösen Verknüpfung, die bis auf die Wortebene reicht. Die Bezüge werden durch Marginalspaltenverweise im Quellenteil und beim Abdruck eines ‚Quellenbezogenen Textes‘ hergestellt und durch typografische Kennzeichen im abgedruckten

⁴⁹ Sahle (Anm. 11), 240 [Hervorhebung im Original].

⁵⁰ Johann W. von Goethe, *Goethes Werke*, hg. im Auftrage der Großherzogin Sophie von Sachsen, 143 Bde. Weimar 1887–1919; Bde. 8, 13 u. 39: *Götz von Berlichingen*, Bde. 14 u. 39: *Faust I*, Bde. 10 u. 39: *Iphigenie auf Tauris*, Bde. 21–23 u. 51/52: *Wilhelm Meisters Lehrjahre*, Weimar 1887–1919.

⁵¹ Johann W. von Goethe, *Werke Goethes*, hg. von der Deutschen Akademie der Wissenschaften zu Berlin, Berlin 1952–1966, abgebrochen, danach weitere Bände als Einzelausgaben, Bd.: *Die Leiden des jungen Werthers*, Berlin 1954.

⁵² Goethe (Anm. 51), Bd.: *Faust*, Ergänzungsband 3, Berlin 1958.

⁵³ Johann W. von Goethe, *Goethes Iphigenie auf Tauris*. In *vierfacher Gestalt*, hg. von Jakob Baechtold, Freiburg i.Br./Tübingen 1883; zweite Ausgabe [als Titeldruck], Freiburg i.Br. 1888.

Quellentext wie im quellenbezogenen Werktextabdruck im Detail sichtbar gemacht.⁵⁴

Die Printeditiionspraxis hat also auch für komplexere, scheinbar das Buchmedium sprengende editorische Darstellungsnotwendigkeiten praktische Lösungen gefunden. Ohne Zweifel hat hier die *Digitale Edition* allerdings grundsätzlich mehr Möglichkeiten anzubieten, weil sie noch mehr Sichten anlegen kann und gleichzeitig nach Benutzerwunsch etwa in verschiedenen Kombinationen aufrufbar bzw. ein- und ausblendbar gestalten kann. Dies hat aber wiederum mit der schon erörterten medial bedingten Funktionalität der *Digitalen Edition* zu tun. Eine individuelle Generierbarkeit durch den Benutzer kann die Printedition in diesem Maße nicht anbieten. Doch auch in der *Digitalen Edition* ist vom Benutzer nur aufrufbar, was zuvor implementiert bzw. kodiert wurde. Das entspricht im Allgemeinen zunächst aber dem interessegeleiteten Abruf von Informationen, die der Benutzer in einer mit reichen Angeboten versehenen Printedition selektiert.

Eines aber ist aus der obigen Betrachtung von Theorie, Methodik und Praxis in Hinblick auf das in Anschlag gebrachte ‚digitale Paradigma‘ festzuhalten: Die Qualitätsinnovation dürfte – zumindest was die neugermanistische Edition betrifft – deutlich geringer sein, als es zunächst den Anschein hat.⁵⁵ Das soll die *Digitale Edition* aber keineswegs abwerten; sie ist allemal die Zukunft des Edierens.⁵⁶ Zudem ist zu bedenken, dass die angeführten Fälle aus dem Bereich der theoretisch und methodisch herausragenden Printeditionen, quasi der neugermanistischen Leuchtturm-Editionen stammen. In der Breite sind die Printeditionen, erst recht auf dem Feld der Studienausgaben, den ausgestellten Paradigmen aber nicht immer unbedingt gefolgt. Das könnte bei der *Digitalen Edition* anders sein, weil die erörterten Paradigmen aufgrund der medialen Bedingungen der *Digitalen Edition* in diese in der ganzen Breite der Ausgabentypen eingehen dürften. Eine besondere Spezifität der *Digitalen Edition* dürfte schließlich auf dem Feld der Nachnutzbarkeit ihrer Daten liegen, die in

⁵⁴ Siehe die Vorstellung des Verfahrens bei Burghard Dedner, „Die Darstellung von Quellenabhängigkeiten anhand von Beispielen“, in: *editio* 11 (1997), 97–115.

⁵⁵ Siehe auch die – allerdings etwas zu pauschal formulierte – Perspektive: „Digital philology [...] has broad new tools, efficient data handling, and a wholly modern outlook to philology. Yet the texts studied and edited remain grossly the same, and the questions asked by digital philologists do not seem to be qualitatively different from the questions asked by pre-digital philologists. [...] It should be added that in many cases digital tools are not new. They are simply more efficient“, Odd E. Haugen/Daniel Apollon, „The Digital Turn in Textual Scholarship. Historical and Typological Perspectives“, in: Daniel Apollon/Claire Belisle/Philippe Regnier (Hg.), *Digital Critical Editions*, Urbana/Chicago/Springfield 2014, 35–57, hier: 50; siehe auch Daniel Apollon/Claire B elisle/Philippe R egnier, „Introduction. As Texts Become Digital“, in: Dies. (Hg.), *Digital Critical Editions*, Urbana/Chicago/Springfield 2014, 1–32, hier: 19, mit weiteren Nachweisen.

⁵⁶ Die folgenden Ausf uhungen in diesem Absatz greifen pr azisierend Anregungen aus der Diskussion dieses Beitrags auf; siehe das zum Beitrag geh orende Diskussionsprotokoll.

ihren je spezifischen Teilen Grundlagen für andere Anwendungsbereiche der digitalen Literaturwissenschaft bereitstellen können, etwa für annotierende Erschließungsverfahren narratologischer, stilistisch-rhetorischer oder formal-strukturaler (für die Lyrik z. B. die Metrik oder das Reimschema betreffender) Interessen. Zudem kann die *Digitale Edition* ihre Daten als Teilmenge kleinerer oder größerer Korpora zur Verfügung stellen, die nach Maßgaben von *Distant-Reading*-Verfahren für quantitative Analysen ausgewertet werden könnten. Zumindest eröffnet die durch das Digitale neu und anders gegebene Interoperabilität der Editionsbestandteile die Möglichkeit für solche Nachnutzungen.⁵⁷ Auf jeden Fall aber kann die *Digitale Edition* gegenüber der Printedition mit der Quantität ihrer Editionseinheiten aufwarten, die auf der Funktionalität ihrer medienspezifischen Modi beruht.

5 Selbstverständnis des Präsentierten: Edition, Archiv, Portal u. a

Die Modi der *Digitalen Edition* hängen nun aber zunächst einmal von den Rahmen und leitenden Charakteristika ab, die sich die Edition gibt. Auf ihnen beruht das Selbstverständnis des Präsentierten. Und es ist innerhalb der editorischen Praxis des Digitalen noch nicht klar, welche Leitcharakteristika für die wissenschaftliche Edition als wesentlich empfunden werden – und vor allem, ob das herkömmliche, in der Geschichte der Printedition entwickelte Verständnis von ‚Edition‘ im neuen Medium überhaupt erhalten bleibt. Das Spektrum der Ausgabentypen ist schon in der Printedition breit, wenn man allein auf die – rasterartige – Typologie ‚Historisch-kritische Ausgabe – Kritische Ausgabe – Studienausgabe – Leseausgabe‘ schaut.⁵⁸ Solche Differenzierungen haben für die *Digitale Edition* bisher allerdings keine Rolle gespielt. Die Diskussion und die Praxis der *Digitalen Edition* beruhen stattdessen auf der Grundsatzfrage der Unterschreitung oder der Überschreitung der bisherigen Rahmen der Printedition.

Bemerkbar wird dies an den Namensgebungen, die schon anzeigen, dass sich im digitalen Medium neue Typen von Überlieferungspräsentationen entwickelt haben. Der maßgeblichste ist der des ‚Archivs‘, der gerade im englischsprachigen Raum gehäuft auftritt. Schon die Namen der Projekte sind sprechend,

⁵⁷Insofern ist die folgende Feststellung zumindest in Hinblick auf die Chancen der Interoperabilität als etwas verkürzend zu bewerten: „the digital revolution essentially is the introduction of more efficient tools. It is not the introduction of new scientific methods or critical ideologies. It is not the introduction of new points of view“, Haugen/Apollon (Anm. 55), 55.

⁵⁸Vgl. Dirk Göttsche, „Ausgabentypen und Ausgabenbenutzer“, in: Rüdiger Nutt-Kofoth/Bodo Plachta/H.T.M. van Vliet u. a. (Hg.), *Text und Edition. Positionen und Perspektiven*, Berlin 2000, 37–63; Bodo Plachta, *Editionswissenschaft. Eine Einführung in Methode und Praxis der Edition neuerer Texte*, 3., ergänzte und aktualisierte Aufl., Stuttgart 2013, 11–26.

z. B. *Shakespeare Electronic Archive*, *Blake Archive*, *Rossetti Archive*, *Dickinson Electronic Archives*, *Shelley-Godwin Archive*, *Beckett Archive*.⁵⁹ Gelegentlich ist eine solche elektronische Sammlung auch nach dem Modell der Bibliothek angedacht worden.⁶⁰ Im Prinzip heißt das Konzept aber *Electronic Archive*.⁶¹ Es kann in seinen Inhalten durchaus denjenigen der Historisch-kritischen Ausgabe ähneln,⁶² sieht aber im Regelfall die Sammlung der Überlieferungsträger als eine Hauptaufgabe an. Gerade bei literarischen Autoren, die auch in anderen Künsten tätig waren (wie bei Blake und Rossetti), ist ein Schwerpunkt auf der visuellen Zurverfügungstellung von Bildern und Handschriftenfotografien verständlich. Das strenge Archiv-Konzept in seiner Differenz zu demjenigen der Historisch-kritischen Ausgabe ist im Übrigen vor einem halben Jahrhundert schon für die neugermanistische Printedition vorgeschlagen worden.⁶³ Im digitalen Bereich liegen auch Mischformen vor, etwa wenn für Nietzsche unter dem Namen

⁵⁹ Siehe William Shakespeare, *Shakespeare Electronic Archive*, hg. von Peter S. Donaldson/Larry Friedlander/Janet H. Murray u. a., Cambridge, MA, <http://shea.mit.edu/>; William Blake, *The William Blake Archive*, hg. von Morris Eaves/Robert N. Essick/Joseph Viscomi, <http://www.blakearchive.org/>; Dante G. Rossetti, *Rossetti Archive*, hg. von Jerome J. McGann, <http://www.rossettiarchive.org/>; Emily Dickinson, *Dickinson Electronic Archives*, hg. von Martha N. Smith/Marta Werner/Jessica Beard u. a., <http://www.emilydickinson.org/>; Percy B. Shelley/Mary Wollstonecraft Shelley/William Godwin/Mary Wollstonecraft, *The Shelley-Godwin Archive*, hg. von Neil Fraistat/Elizabeth Denlinger/Raffaele Vigilanti u. a., <http://shelleygodwinarchive.org/> (letzter Aufruf der Links 22.9.2017). Beim *Beckett*-Projekt findet sich die ‚Archiv‘-Bezeichnung nur in der URL, nicht im Projektnamen: Samuel Beckett, *Samuel Beckett Digital Manuscript Project*, hg. von Dirk Van Hulle/Mark Nixon/Vincent Neyt, Brüssel 2011 ff., <http://www.beckettarchive.org/> (letzter Aufruf 22.9.2017).

⁶⁰ „This vision of an electronic scholarly edition begins to resemble an archive of editions with annotations, contexts, parallel texts, reviews, criticism, and bibliographies of reception and criticism. In effect it is really a library we want“, Peter Shillingsburg, „Principles for Electronic Archives, Scholarly Editions, and Tutorials, in: Richard J. Finneran (Hg.), *The Literary Text in the Digital Age*, Ann Arbor 1996, 23–35, hier: 24.

⁶¹ Peter L. Shillingsburg, *From Gutenberg to Google. Electronic Representations of Literary Texts*, Cambridge, UK/New York u. a. 2006, 148.

⁶² Siehe etwa ebd., 48: „[...] beginning with images of original editions in contemporary bindings, and then including digitized texts, and later expanding to include textual, production, and other historical annotations, as well as texts newly edited by bold editors.“

⁶³ Klaus Kanzog, *Prolegomena zu einer historisch-kritischen Ausgabe der Werke Heinrich von Kleists. Theorie und Praxis einer modernen Klassiker-Edition*, München 1970, 15–29 (Kapitel „Die Archiv-Ausgabe“ und „Die historisch-kritische Ausgabe“); wiederabgedruckt in Nutt-Kofoth (Anm. 26), 335–344; zur Aufnahme von Kanzogs Vorschlag für die Printedition siehe Hans Zeller, „Die Faksimile-Ausgabe als Grundlagenedition für Philologie und Textgenetik. Ein Vorschlag“, in: Hans Zeller/Gunter Martens (Hg.), *Textgenetische Edition*, Tübingen 1998, 80–100; editorisch realisiert z. B. in der Printausgabe Georg Büchner, *Woyzeck. Faksimileausgabe der Handschriften*, bearbeitet von Gerhard Schmid, Faksimile/Transkription, Wiesbaden 1981.

Nietzsche Source neu hergestellte digitale Faksimiles neben der retrodigitalisierten historisch-kritischen Printausgabe zur Verfügung gestellt werden.⁶⁴

Der Name ‚Source‘ umgeht die Bezeichnungen ‚Archiv‘ und ‚Edition‘ möglicherweise, um den Verbund beider Charakteristika anzuzeigen. Er hat sich allerdings ansonsten nicht durchgesetzt. Wie die digitale Präsentation edierten Materials zu bezeichnen ist, ist in der Tat noch völlig unregelt. Für die Zusammenführung von Material wird auch der Name ‚Portal‘ verwendet, wie etwa beim *Heinrich-Heine-Portal*, das zwei historisch-kritische Heine-Printausgaben verbindet und mit Digitalisaten der Textträger anreichert.⁶⁵ Das *Georg-Büchner-Portal* wiederum versammelt Informationen, Dokumente und Aufsätze zu Büchner auf Grundlage der historisch-kritischen Marburger *Büchner-Ausgabe* und kündigt zugleich an, diese Printedition zu retrodigitalisieren und in das Portal einzubinden.⁶⁶ Werkmaterialien in Kombination mit Forschungsbeiträgen legt dagegen das Projekt *Handkeonline* unter der Bezeichnung ‚Plattform‘ ab.⁶⁷ Gerade in Hinblick auf die Einbindung unterschiedlicher Kontextmaterialien sind im anglo-amerikanischen Raum für die *Digitale Edition* auch noch ganz andere Namen vorgeschlagen worden, neben den Bezeichnungen ‚database‘, ‚project‘ oder ‚thematic research collection‘ etwa ‚arsenal‘⁶⁸ oder ‚knowledge site‘.⁶⁹ Während ‚database‘ deshalb problematisch ist, weil die *Digitale Edition* im Ganzen nicht

⁶⁴ <http://www.nietzschesource.org/>: Friedrich Nietzsche, *Digitale Faksimile-Gesamtausgabe. Nach den Originalmanuskripten und Originaldrucken der Bestände der Klassik Stiftung Weimar*, hg. von Paolo D'Iorio, Nietzsche Source, Paris 2009 ff., <http://www.nietzschesource.org/DFGA/> (letzter Aufruf 22.9.2017); Friedrich Nietzsche, *Digitale Kritische Gesamtausgabe. Werke und Briefe, Nietzsche Source*, hg. von Paolo D'Iorio, Paris 2009 ff., Berlin/New York, 1967 ff., <http://www.nietzschesource.org/#eKGWB> (letzter Aufruf 22.9.2017); als Printausgabe: Friedrich Nietzsche, *Werke. Kritische Gesamtausgabe*, begründet von Giorgio Colli/Mazzino Montinari, weitergeführt von Wolfgang Müller-Lauter/Karl Pestalozzi, Berlin u. a. 1967 ff.

⁶⁵ Heinrich Heine, *Heinrich-Heine-Portal*, hg. von Thomas Burch/Joseph A. Kruse/Bernd Füllner u. a. 2002–2009, <http://www.heine-portal.de/> (letzter Aufruf 22.9.2017). Bei den Printausgaben handelt es sich um Heinrich Heine, *Säkularausgabe. Werke, Briefwechsel, Lebenszeugnisse*, hg. von den Nationalen Forschungs- und Gedenkstätten der klassischen deutschen Literatur in Weimar [seit 1993: Stiftung Weimarer Klassik, seit 2008: Klassik Stiftung Weimar] und dem Centre National de la Recherche Scientifique in Paris, Berlin/Paris 1970 ff., und Heinrich Heine, *Historisch-kritische Gesamtausgabe der Werke*, hg. von Manfred Windfuhr im Auftrag der Landeshauptstadt Düsseldorf, 16 Bde. in 23, Hamburg 1973–1997.

⁶⁶ Siehe *Georg-Büchner-Portal*, hg. von Burghard Dedner, Marburg, <http://buechnerportal.de/> (letzter Aufruf 22.9.2017); Büchner (Anm. 45).

⁶⁷ Peter Handke, *Handkeonline*, hg. von Klaus Kastberger/Katharina Pektor/Christoph Kepplinger-Prinz, <http://handkeonline.onb.ac.at/> (letzter Aufruf 22.9.2017). Siehe auch Klaus Kastberger/Christoph Kepplinger-Prinz, „Handkeonline. Eine Forschungsplattform zu Peter Handke“, in: *editio* 27 (2013), 205–215.

⁶⁸ Die letzte Bezeichnung ist in Abwägung anderer bevorzugt bei Kenneth M. Price, „Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?“, in: *Digital Humanities Quarterly* 3/3 (2009), Kap. 40, <http://www.digitalhumanities.org/dhq/vol/3/3/000053/000053.html> (letzter Aufruf 22.9.2017).

⁶⁹ Vorgeschlagen von Shillingsburg (Anm. 61), 5, 97 u. ö.

nur aus ihren Daten besteht, sondern auch aus ihrer Präsentationsoberfläche, über die die Daten für den Benutzer erst zugänglich werden, sind ‚project‘ und ‚thematic research collection‘ so allgemein, dass sie kaum als markante und spezifische Bezeichnung für die *Digitale Edition* nutzbar sind. Für ‚arsenal‘ dürfte dasselbe gelten. Peter Shillingsburgs Vorschlag ‚knowledge site‘ ist dagegen des Öfteren – skeptisch oder zustimmend – in der editorischen Diskussion aufgegriffen worden.⁷⁰ Doch kann auch er nicht wirklich den Charakter der *Digitalen Edition* bezeichnen. Eine ‚Wissensstätte‘ oder ein ‚Wissensspeicher‘ war ja auch schon die Historisch-kritische Printedition; und die ‚knowledge site‘ soll ja auch explizit nicht-editorische Textsorten – wie Interpretationen – enthalten.⁷¹ Insofern geht es primär um eine dem Inhalt der Edition adäquate Benennung.

Solange das Format ‚Edition‘ als wissenschaftliches Produkt im Mittelpunkt steht, muss es keinen Streit um die Ausbildung synonym gebrauchter Bezeichnungen geben, wie sie im englischsprachigen Raum auftauchen, etwa *Digital Scholarly Edition*⁷² und *Digital Critical Edition*.⁷³ Die Edition sollte stattdessen jenen Ort selbst markieren, an dem sie innerhalb von digitalen Wissensrepräsentationen steht, d. h. hinsichtlich dessen, was sie enthält und wessen Teil sie ist. So können Editionen zwar als ‚virtuelle Archive‘ verstanden werden,⁷⁴ doch wird man überlegen müssen, warum sich *Digitale Editionen* vielfach der Archivfunktion annähern.⁷⁵ Weil die Funktionalität des Mediums es erlaubt, Digitalisate als Abbildungen der Textträger in großer Menge in die Edition einzubinden (quantitativer Mehrwert gegenüber der Printedition), heißt dies ja noch nicht, dass sich die *Digitale Edition* darin erschöpfen sollte oder dies zu ihrem Hauptmerkmal erklärt. Ganz in diesem Sinne hat Peter Robinson 2013 den Unterschied zwischen einem *Digitalen Archiv* und einer *Digitalen Edition* gekennzeichnet:

⁷⁰ Philippe Régnier, „Ongoing Challenges for Digital Critical Editions“, in: Ders./Apollon/Belisle (Anm. 55), 58–80, hier: 68; Dirk Van Hulle, „Modelling a Digital Scholarly Edition for Genetic Criticism: A Rapprochement“, in: *Variants* 12/13 (2016 [2017]), 34–56, hier: 37 bzw. online Kap. 5, <http://variants.revues.org/293> (letzter Aufruf 22.9.2017).

⁷¹ Vgl. Thomas Stäcker, „Creating the Knowledge Site – elektronische Editionen als Aufgabe einer Forschungsbibliothek“, in: *Bibliothek und Wissenschaft* 44 (2011), 107–126, hier: 112.

⁷² Siehe Pierazzo (Anm. 3), 193.

⁷³ Siehe nur den Titel von Apollon u. a. (Anm. 55).

⁷⁴ Siehe Peter Shillingsburg, „Development Principles for Virtual Archives and Editions“, in: *Variants* 11 (2014), 11–28.

⁷⁵ Vgl. auch Patrick Sahle, „Digitales Archiv – Digitale Edition. Anmerkungen zur Begriffsklärung“, in: Michael Stolz/Lucas Marco Gisi/Jan Loop (Hg.), *Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien. Eine Standortbestimmung*, Zürich 2007, 64–84; Wolfgang Lukas, „Was ist das Digitalisierungsinteresse der geistes- und kulturwissenschaftlichen Forschung?“, in: LVR-Archivberatungs- und Fortbildungszentrum (Hg.), *Digital und analog. Die beiden Archiwelten*. 46. Rheinischer Archivtag. Ratingen 21.–22. Juni 2012. Beiträge, Bonn 2013, 32–47, hier: 38.

[T]he common first error made by digital humanists, – and, I fear, by many editors: that the needs of the textual scholars, and indeed the interests of readers, can be perfectly served by digital archives. They cannot. An edition is an argument about a text. We need arguments; without arguments, our archives are inert bags of words and images.⁷⁶

Insofern darf ein kritischer Blick zu Recht auf der annoncierten Präferenz von vor allem dokumentarisch-faksimilierenden – und damit dem Archivgedanken Vorrang einräumenden – Editionsmodellen liegen.⁷⁷

6 Reformulierung des neugermanistischen Editionstypus: Die *Digitale Historisch-kritische Ausgabe* im Wissensverbund

Vorgeschlagen werden soll hier ein Editionstypus, der im digitalen Medium nicht hinter die editorischen Standards der Historisch-kritischen Printausgabe zurückfällt und zugleich die Möglichkeiten von Wissensordnungen und Wissenspräsentationen im digitalen Medium nutzt. Die in den theoretischen Diskussionen immer wieder durchscheinende Charakterisierung der Historisch-kritischen Ausgabe als eines dem digitalen Medium inadäquaten Modells, das sich letztlich schon im 19. Jahrhundert entwickelt hat, sollte vor dem Hintergrund der neugermanistischen Editions-geschichte des letzten halben Jahrhunderts als grobe Verzeichnung, ja als Fehldeutung erkennbar geworden sein. Insofern ist es ganz in diesem Sinne wegweisend, dass sich – nach Diskussionen – die digitale Ausgabe von Goethes *Faust* (3. Beta-Version 2017) explizit *Historisch-kritische Edition* nennt.⁷⁸

In Hinblick auf die Position der *Digitalen Historisch-kritischen Ausgabe* (DHKA) in der digitalen Wissensdarbietung ließe sich folgende Ordnung anbieten: Primär archivalisch angelegte oder dokumentenzentrierte Ausgaben bilden aus

⁷⁶Peter Robinson, *Why digital humanists should get out of textual scholarship / What digital humanists don't know about scholarly editing: what scholarly editors don't know about the digital world*, http://www.academia.edu/4124828/SDSE_2013_why_digital_humanists_should_get_out_of_textual_scholarship (letzter Aufruf 22.9.2017).

⁷⁷Siehe zur letzteren Position Hans W. Gabler, „Das wissenschaftliche Edieren als Funktion der Dokumente“, in: *Jahrbuch für Computerphilologie* 8 (2006), <http://computerphilologie.digital-humanities.de/jg06/gabler.html> (letzter Aufruf 22.9.2017); die Kritik daran bei Peter Robinson, „Towards a Theory of Digital Editions“, in: *Variants* 10 (2013), 105–131, hier: bes. 126 f.

⁷⁸Johann W. von Goethe, *Faust. Historisch-kritische Edition*, hg. von Anne Bohnenkamp/Silke Henke/Fotis Jannidis u. a., Beta-Version 3, Frankfurt a. M./Weimar/Würzburg 2016/17, <http://beta.faustedition.net/> (letzter Aufruf 22.9.2017). Die Ausgabe versteht sich zugleich als Modell für die *Digitale Historisch-kritische Ausgabe*, siehe Anne Bohnenkamp/Silke Henke/Fotis Jannidis u. a., „Die digitale ‚Faust‘-Edition. Zur neuen historisch-kritischen Ausgabe von Goethes Drama“, in: *Goethe-Jahrbuch* 133, 2016 [2017], 150–162, hier: 162: „Erklärtes Ziel der neuen *Faust*-Edition ist [...] die Entwicklung eines Modells für historisch-kritische Ausgaben im elektronischen Medium.“

struktureller Perspektive nur ein Teilelement der *DHKA*.⁷⁹ Sie sind dann als Bestandteil jenes Elementenbaukastens des historisch-kritischen Ausgaben-typs zu verstehen, wie er schon für die Printedition ausdifferenziert wurde.⁸⁰ Eng zusammen gehörende, an sich selbstständige Editionen können als *DHKA* wiederum in einer größeren Umgebung zusammengebunden werden, wie es etwa das *Proypläen*-Projekt für die in Arbeit befindlichen verschiedenen Teilausgaben zu Goethes *Biographica* plant.⁸¹ Darüber hinaus kann die *DHKA* selbst ein Teilelement umfangreicherer, nicht allein editorisch ausgerichteter digitaler ‚Wissensstätten‘ bilden, etwa wenn sie in ein Portal, eine Plattform, eine *Knowledge Site*, die durch welchen Rahmen auch immer konturiert sind, inkorporiert wird. Insofern kann die *DHKA* in einer zunehmend digitalen Literaturwissenschaft leichter zu anderen literaturwissenschaftlichen Themenfeldern – etwa im Sinne eines ‚Dokumentknoten[s]‘⁸² – in ein Verhältnis gesetzt werden, z. B. zu interpretativen oder literaturgeschichtlichen Projekten, was die mangelhafte Nutzung kritischer Ausgaben in der printgestützten Literaturwissenschaft schrittweise zu beheben geeignet sein könnte.⁸³ Schon vor mehr als zwei Jahrzehnten ist in ähnlichem Sinne das Konzept einer literarischen Epochendatenbank formuliert

⁷⁹Das lasse sich auch auf Textsortenfragen übertragen. Insofern würden auch etwa Notizbuch-Editionen als Teilelement einer *Digitalen Historisch-kritischen Gesamterkaufgabe* fungieren. Falls es eine solche nicht gibt, können sie selbständig erscheinen; das aktuell wohl ambitionierteste Projekt ist Theodor Fontane, *Notizbücher. Digitale genetisch-kritische und kommentierte Edition*, hg. von Gabriele Radecke, <https://fontane-nb.dariah.eu/index.html> (Version 0.1 vom 7. Dezember 2015, im Folgenden weiter aktualisiert; letzter Aufruf 22.9.2017). Siehe dazu Gabriele Radecke/Heike Neuroth/Martin de la Iglesia u. a., „Vom Nutzen digitaler Editionen. Die Genetisch-kritische Hybrid-Edition von Theodor Fontanes Notizbüchern erstellt mit der Virtuellen Forschungsumgebung TextGrid“, in: Thomas Bein (Hg.), *Vom Nutzen der Editionen. Zur Bedeutung moderner Editorik für die Erforschung von Literatur- und Kulturgeschichte*, Berlin/Boston 2015, 277–295.

⁸⁰Siehe Hans Zeller, „Für eine historische Edition. Zu Textkonstitution und Kommentar“, in: Georg Stötzel (Hg.), *Germanistik – Forschungsstand und Perspektiven. Vorträge des Deutschen Germanistentages 1984*, 2. Teil: *Ältere Deutsche Literatur; Neuere Deutsche Literatur*, Berlin/New York 1985, 305–323, hier: 320–323; Elisabeth Höpker-Herberg/Hans Zeller, „Der Kommentar, ein integraler Bestandteil der historisch-kritischen Ausgabe?“, in: *editio* 7 (1993), 51–61.

⁸¹Zu Konzept und Ziel siehe Johann W. von Goethe, *Proypläen. Goethes Biographica*, hg. von Bernhard Fischer/Klaus Manger/Ernst Osterkamp, Weimar/Frankfurt a. M. 2015 ff., <http://www.goethe-biographica.de/> (letzter Aufruf 22.9.2017).

⁸²Stäcker (Anm. 71), 108: „Wenn die Bibliothek der Zukunft eine Bibliothek elektronischer Texte ist, dann ist die elektronische oder digitale Edition nichts anderes als ein in einem textuellen Zusammenhang einer Bibliothek stehender besonderer Text oder ein Dokumentknoten, der dazu dient, weitere Dokumente miteinander zu verbinden.“

⁸³Zur Problematik s. Rüdiger Nutt-Kofoth, „Wie werden neugermanistische (historisch-)kritische Editionen für die literaturwissenschaftliche Interpretation genutzt? Versuch einer Annäherung aufgrund einer Auswertung germanistischer Periodika“, in: Bein (Anm. 79), 233–245.

worden, deren einer Teil aus einer *Digitalen Edition* bestehen würde.⁸⁴ Allerdings sollte die *DHKA* in ihren Umgebungen – als untergeordnetes Teilelement einer digitalen ‚Wissensstätte‘ oder als selbstständiges Element innerhalb der allgemeinen literaturwissenschaftlichen digitalen Ordnung – immer klar identifizierbar sein. Nur so ist auch für den Benutzer auf den ersten Blick ersichtlich, wo die Qualitätsstandards der Historisch-kritischen Ausgabe – skrupulöse Textkritik für jedwede Textdarstellung und Vollständigkeit der Textgenese sowie aller entstehungsgeschichtlichen wie die Historizität des Editionsobjekts betreffenden Sachverhalte – in einer digitalen ‚Wissensstätte‘ auf jeden Fall gültig sind. Das ist auch deshalb von besonderem Belang, weil unter *Digitalen Editionen* gerade im Teilbereich Briefedition solche Projekte auf den Weg gebracht sind, die z. B. das Kriterium der ‚kritischen Textkonstitution‘ zugunsten anderer Interessen, etwa Briefnetzwerkdarstellung oder Textmengenerfassung, zurückstellen.⁸⁵ Insofern sollte die *DHKA* – wenn sie in ihrer vollen Breite realisiert wird – archivalisch-dokumentierende von textkonstituierenden (im Sinne des statischen wie auch des dynamischen Textverständnisses) und diese wiederum von kommentierend-erschließenden Teilen abgrenzen und all diese im Sinne des für die Printedition entwickelten Baukastensystems (s. o.) jeweils deutlich ausweisen. Dadurch wären nicht nur innerhalb der Spannweite der Edition und zwischen den Polen von Befund und Deutung die je einzelnen Bestandteile differenziert verortet, sondern die Edition könnte den in jüngeren Historisch-kritischen Printausgaben häufiger festzustellenden Verzicht auf Kommentar und Erläuterungen wieder aufheben.⁸⁶ So könnten gerade über die erschließenden Editionsanteile Vermittlungsbrücken zu anderen, stärker interpretativen Anteilen von digitalen ‚Wissensstätten‘, deren Teil die *DHKA* bilden oder mit denen sie verknüpft werden kann, gebaut werden; Brücken, die Editionswissenschaft und Literaturwissenschaft wieder stärker

⁸⁴ Walter Fanta, „Die Computer-Edition des Musil-Nachlasses. Baustein einer Epochendatenbank der Moderne“, in: *editio* 8 (1994), 127–157.

⁸⁵ Siehe etwa das Exilbriefnetz-Projekt *Vernetzte Korrespondenzen*, Projektleitung: Thomas Burch u. a., <http://exilnetz33.de> (letzter Aufruf 22.9.2017) oder August W. Schlegel, *Digitale Edition der Korrespondenz August Wilhelm Schlegels*, Projektleitung: Jochen Strobel, <http://august-wilhelm-schlegel.de/briefedigital/>, Beta-Version, Stand der Daten: 4.7.2017 (letzter Aufruf 22.9.2017). Vgl. in diesem Zusammenhang Rüdiger Nutt-Kofoth, „Briefe herausgeben: Digitale Plattformen für Editionswissenschaftler und die Grundfragen der Briefedition“, in: Kristina Richts/Peter Stadler (Hg.), „*Ei, dem alten Herrn zoll' ich Achtung gern*“. *Festschrift für Joachim Veit zum 60. Geburtstag*, München 2016, 575–586, hier: bes. 583–586. Eine Mahnung in Hinblick auf die geringere Wertschätzung der Textkritik bei größeren Digitalisierungsprojekten auch bei Bodo Plachta, „Edition und Bibliothek“, in: *Bibliothek und Wissenschaft* 44 (2011), 23–36, hier: 34.

⁸⁶ Das Gegenmodell eines umfangreichen, äußerst detaillierten Kommentars findet sich allerdings in jüngeren Print-HKAs auch; siehe etwa die *Büchner-Ausgabe* (Anm. 45).

zusammenführen könnten,⁸⁷ weil diese im digitalen Medium ihre jeweiligen Leistungen als abgegrenzt-gestaffelte, aber enger aufeinander zu beziehende – und dafür digital verlinkbare, somit benutzerseitig leichter als ein größerer Zusammenhang wahrnehmbare – verstehen könnten.

In Anknüpfung an die Standards der Printedition und in Nutzung der Möglichkeiten des digitalen Mediums enthält die *DHKA* folgende Elemente:

- vollständige Präsentation aller relevanten Textträger als Digitalisate;
- vollständige diplomatische Transkription aller Handschriften; im Ideal auch vollständige zeilengetreue Transkription aller Drucke; beide als Paralleldarstellung zu den Digitalisaten anzeigbar, die Handschriftentranskription im Idealfall auch zusätzlich positionsgenau im Digitalisat als überlagernde Darstellung ein- und ausblendbar;
- vollständige Darstellung der Textgenese, sodass geringe wie umfangreiche Änderungsprozesse sichtbar werden;
- konstituierte Texte von Fassungen oder Textstadien, soweit dies editorisch angemessen ist;
- ausführliche Beschreibung der Textträger (Metadaten) einschließlich möglicher grafischer Darstellungen (etwa Lagenbeschreibung bei Handschriften) und textträgerkontextualisierender Informationen (etwa druckanalytischer Fragen, je nach Sachlage auch typografiekritischer Aufarbeitungen bei Drucken);⁸⁸
- Präsentation sämtlicher Dokumente zu Entstehung und zeitgenössischer Rezeption (zu Lebzeiten des Autors);
- ausführliche Entstehungs- und zeitgenössische Rezeptionsgeschichte in Auswertung der Dokumente zur Entstehung;⁸⁹

⁸⁷ Siehe die Feststellung, dass „Literaturwissenschaft und Editionsphilologie [...] sich nicht mehr viel zu sagen haben“, bei Klaus Hurlebusch, *Klopstock, Hamann und Herder als Wegbereiter autorzentrischen Schreibens. Ein philologischer Beitrag zur Charakterisierung der literarischen Moderne*, Tübingen 2001, 69.

⁸⁸ Die analytische Druckforschung ist editorisch allemal vernachlässigt; zur Relevanz siehe Martin Boghardt, *Analytische Druckforschung. Ein methodischer Beitrag zu Buchkunde und Textkritik*, Hamburg 1977 und Martin Boghardt, *Archäologie des gedruckten Buches*, hg. von Paul Needham/Julie Boghardt, Wiesbaden 2008. Eine typografiekritische editorische Aufarbeitung mit Zielrichtung auf eine *Digitale Edition* ist entworfen bei Frederike Neuber, „Typografie und Varianz in Stefan Georges Werk. Konzeptionelle Überlegungen zu einer ‚typografiekritischen‘ Edition“, in: *editio* 31 (2017), 205–232, hier: bes. 228–231.

⁸⁹ Eine solche heuristisch bedeutsame Trennung zwischen Dokumentenlistung und -auswertung im Sinne der Trennung von ‚Befund‘ und ‚Deutung‘ ist bisher nur in jüngeren Bänden der *Schiller-Nationalausgabe* vorgenommen worden; siehe etwa Friedrich Schiller, *Schillers Werke. Nationalausgabe*, 1940 begründet von Julius Petersen, fortgeführt von Lieselotte Blumenthal, Benno von Wiese, Siegfried Seidel, hg. von Norbert Oellers im Auftrag der Stiftung Weimarer Klassik und des Schiller-Nationalmuseums in Marbach, Bd. 5, Neue Ausgabe: *Kabale und Liebe, Semele, Der versöhnte Menschenfeind, Körners Vormittag*, hg. von Herbert Kraft/Claudia Pilling/Gert Vonhoff, Weimar 2000.

- Volltextdarstellung aller Quellen und ihrer Bezugsintensität, eventuell mit Digitalisaten der Textträger;
- optional: Überblickskommentar und Einzelstellenerläuterungen; hier wären wiederum Referenzen auf historische und literarische Folien, direkte und indirekte (bzw. auch markierte und nicht markierte) Zitate, Anspielungen etc. durch Verlinkungen idealerweise auf die Digitalisate von zeitgenössischen bzw. – falls nachweisbar – vom Autor benutzten Drucken ein Gewinn.

Den besonderen Mehrwert der *Digitalen Edition* bietet zum einen die intensive Verlinkung, zunächst einmal vor allem die editionsinterne, die Bezüge – etwa zwischen der Überlieferungsbeschreibung und den Digitalisaten, den verschiedenen parallel oder separat aufrufbaren Textdarstellungen oder auch zwischen verschiedenen Fassungen mit Varianzmarkierungen – benutzerseitig sofort aufrufbar macht. Insofern erlaubt die *DHKA* zum anderen vom Benutzer nach seinen Interessen steuerbare Sichten auf die unterschiedlichen Textdarstellungen und Fassungen, wie sie in diesem Umfang in der Printedition nicht erzeugt werden können.⁹⁰ Zudem ist es ein Zugewinn, wenn der archivalische Teil der *DHKA* dem Benutzer durch hochauflösende, zoombare Digitalisate noch eher als die Printedition für viele – keineswegs natürlich für sämtliche – Fragen den Weg ins Archiv ersparen kann, ja die Handschrift gelegentlich besser als am Original lesbar machen kann (und in der Folge zu einer materiellen Schonung der Originale führen kann).

Auf der anderen Seite ist bei den erst langsam an die Öffentlichkeit tretenden neugermanistischen *Digitalen Editionen* genauer zu betrachten, wie sie mit einem Herzstück der neugermanistischen Printedition, der präzisen, minutiösen textgenetischen Darstellung, umgehen. Während für den eher archivalischen Teil der *Digitalen Edition*, Faksimile und Transkription, ausgefeilte Präsentationen entwickelt worden sind, bleibt bei der Textgenese ein Mangel gegenüber den in der Printedition genutzten Verfahren festzustellen. Denn auch in den jüngsten Werkeditionen zu Goethes *Faust* oder Koeppens *Jugend* werden Einzelstellenänderungen zwar – unterschiedlich – sichtbar gemacht, doch die Darstellung zusammenhängender Änderungskomplexe bleibt un(ter)repräsentiert.⁹¹ Solche Stellen der Mikrogenese bieten noch eine Herausforderung für die *Digitale Edition*, während Elemente der Makrogenese besonders gut etwa durch verlinkte

⁹⁰Vgl. Fotis Jannidis, „Elektronische Edition“, in: Rüdiger Nutt-Kofoth/Bodo Plachta (Hg.), *Editionen zu deutschsprachigen Autoren als Spiegel der Editionsgeschichte*, Tübingen 2005, 457–470, hier: 457; Rüdiger Nutt-Kofoth, „Sichten – Perspektiven auf Text“, in: Anne Bohnenkamp (Hg.), *Medienwandel / Medienwechsel in der Editionswissenschaft*, Berlin/Boston 2013, 19–29. Siehe auch Fotis Jannidis, „Digitale Editionen“, in: *Mitteilungen des Deutschen Germanistenverbandes* 55/3 (2008), 317–332.

⁹¹Goethe (Anm. 78); Wolfgang Koeppen, *Jugend. Textgenetische Edition*, hg. von Katharina Krüger/Elisabetta Mengaldo/Eckhard Schumacher, 2017, <http://www.koeppen-jugend.de> (letzter Aufruf 22.9.2017).

grafische Übersichten – auch im interaktiven Format wie im Fußteil der Textdarstellungsseiten innerhalb der Edition von Hermann Burgers *Lokalbericht*⁹² – wiedergegeben werden können. Außerhalb des Bereichs der Neugermanistik, nämlich in der Antwerpener *Beckett-Edition*, findet sich auch eine animierte grafische Darstellung des Niederschriftverlaufs, deren einzelne Farbpunkte jeweils einen einzelnen Satz repräsentieren.⁹³

Gerade die Textgenese also stellt noch Aufgaben an die *Digitale Edition*. Es verwundert daher auch nicht, dass die *Digitale Werkausgabe* bei der Entwicklung der neugermanistischen Edition – trotz der Pilotprojekte etwa des *Jungen Goethe*, der elektronischen Komponente zur *Keller-Ausgabe* oder der digitalen *Musil-Ausgabe*⁹⁴ – nicht nachdrücklich in den Vordergrund trat, sondern zuletzt vor allem die *Digitale Briefedition* ambitionierte Projekte entwickelt hat, etwa zum Briefnetz der exilierten Autoren der NS-Zeit, zu den August-Wilhelm-Schlegel-Briefen oder zum Gelehrten-Briefwechsel Sauer – Seuffert,⁹⁵ die sich wiederum an wegweisenden nicht-germanistischen digitalen Briefeditionen wie denjenigen zu van

⁹² Hermann Burger, *Lokalbericht. Digitale Edition*, hg. von Peter Dängeli/Magnus Wieland/Irmgard M. Wirtz u. a., Beta-Version, Bern 2016, <http://www.lokalbericht.ch/> (letzter Aufruf 22.9.2017).

⁹³ Beckett (Anm. 59), hier: <http://www.beckettarchive.org/writingsequenceoffinnomtable.jsp> (letzter Aufruf 22.9.2017). Zur Frage der textgenetischen Darstellung in Print- und *Digitalen Editionen* siehe Rüdiger Nutt-Kofoth, „Textgenese analog und digital: Ziele, Standards, Probleme“, in: Anke Bosse/Walter Fanta (Hg.), *Textgenese im digitalen Medium*, Berlin/Boston 2019, 1–19. Die in der Entstehung befindliche *Digitale Schnitzler-Werkausgabe* versteht die textgenetische Darstellung als eine ihrer Hauptaufgaben; siehe dazu Thomas Burch/Stefan Büdenbender/Kristina Fink u. a., „Text[ge]schichten. Herausforderungen textgenetischen Edierens bei Arthur Schnitzler“, in: Katharina Krüger/Elisabetta Mengaldo/Eckhard Schumacher (Hg.), *Textgenese und digitales Edieren. Wolfgang Koeppens ‚Jugend‘ im Kontext der Editionsphilologie*, Berlin/Boston 2016, 87–105; Kathrin Nühlen/Jonas Wolf, „Ansichtssache. Möglichkeiten der Darstellung und Interpretation textgenetischer Varianz am Beispiel der historisch-kritischen Edition ‚Arthur Schnitzler digital‘“, in: Wernfried Hofmeister/Andrea Hofmeister-Winter (Hg.), *Textrevisionen. Beiträge der Internationalen Fachtagung der Arbeitsgemeinschaft für germanistische Edition, Graz, 17.–20. Februar 2016*, Berlin/Boston 2017, 185–198.

⁹⁴ Johann W. von Goethe, *Der junge Goethe in seiner Zeit. Texte und Kontexte*, 2 Bde. und CD-ROM, hg. von Karl Eibl/Fotis Jannidis/Marianne Willems, Frankfurt a. M./Leipzig 1998; Gottfried Keller, *Sämtliche Werke. Historisch-Kritische Ausgabe*, hg. unter der Leitung von Walter Morgenthaler im Auftrag der Stiftung Historisch-Kritische Gottfried Keller-Ausgabe, Basel/Frankfurt a. M./Zürich 1996–2013, digitaler Teil auf beigegebenen CD-ROMs und <https://www.ehkka.ch/ehkka/> (letzter Aufruf 22.9.2017); Robert Musil, *Der literarische Nachlaß*, hg. von Friedbert Aspetsberger/Karl Eibl/Adolf Frisé, Reinbek 1992 [CD-ROM]; Robert Musil, *Klagenfurter Ausgabe. Kommentierte digitale Edition sämtlicher Werke, Briefe und nachgelassener Schriften. Mit Transkriptionen und Faksimiles aller Handschriften*, hg. von Walter Fanta/Klaus Amann/Karl Corino, Klagenfurt, DVD-Version 2009.

⁹⁵ Vernetzte Korrespondenzen (Anm. 85), Schlegel (Anm. 85); Alfred Sauer/Bernhard Seuffert, *Briefwechsel*, hg. von Bernhard Fetz/Hans H. Müller/Desiree Hebenstreit u. a., Wien/Hamburg 2016, <http://sauer-seuffert.onb.ac.at/> (letzter Aufruf 22.9.2017).

Gogh, Escher oder Weber orientieren konnten.⁹⁶ Das dürfte vor allem mit der gegenüber Werken eher geringeren Problemlage in Hinblick auf Textkritik und Textgenese (und damit den Zentren der Historisch-kritischen Ausgabe) zu tun haben, die sich aus der Medienspezifik und den Überlieferungsgegebenheiten des Briefes herleitet, sodass Briefe in dieser Hinsicht editorisch etwas einfacher zu behandeln sind.

7 Postskript

Ob die *DHKA* von einer Printedition, die Teile der *Digitalen Edition* enthält, begleitet wird oder nicht, ist keine prinzipielle Frage, denn die strukturierten Daten der *Digitalen Edition* erlauben allemal Druckderivate auch zu späterer Zeit. De facto sind die allermeisten der oben genannten neugermanistischen *Digitalen Editionen* (etwa *Faust*, *Jugend*, *Lokalbericht*) als Hybridausgaben angelegt. Trotz aller Bemühungen um die Datenauszeichnung in nicht-proprietären Umgebungen und mithilfe zunehmend standardisierter Verfahren – Vorreiter bildet derzeit die *Text Encoding Initiative (TEI)*⁹⁷ – ist letztlich nicht sicher, wie lange eine *Digitale Edition* einschließlich all der in ihr angelegten Funktionalitäten tatsächlich zugänglich bleibt. Das Schicksal der digitalen *Musil-Ausgaben* von 1992 und 2009,⁹⁸ die beide – zuvorderst aufgrund ihrer Bindung an proprietäre, seit längerem nicht mehr weiterentwickelte Software (*FolioViews*) – auf gegenwärtig aktuellen Computern bzw. Betriebssystemen, neuerdings auch aufgrund ihrer Hardware-Bedingungen in Zeiten, in denen Rechner ohne CD-Laufwerke hergestellt werden, nicht mehr bzw. nur mit Umständen gängig sind, ist immerhin ein bedenkliches Beispiel.

An der Frage der Langzeitarchivierung und vor allem auch -gangbarhaltung wird zwar intensiv gearbeitet, doch bleiben Grundsatzfragen ein Problem. So wird die Trennung von Daten und Präsentation der *Digitalen Edition* als Vorteil für die Transmedialisierung begreifbar.⁹⁹ Dies führt jedoch auch dazu, dass sich die Langzeitarchivierung nur auf den Datenkern, nicht aber die Präsentation, also

⁹⁶Vincent van Gogh, *The Letters*, hg. von Leo Jansen/Hans Luijten/Nienke Bakker, Amsterdam/Den Haag 2009 (Version: April 2017), <http://vangoghletters.org> (letzter Aufruf 22.9.2017); Alfred Escher, *Digitale Briefedition*, hg. von Joseph Jung, Launch Juli 2015 (laufend aktualisiert), Zürich, <https://www.briefedition.alfred-escher.ch/> (letzter Aufruf 22.9.2017); Briefabteilung innerhalb von Carl Maria von Weber, *Gesamtausgabe. Digitale Edition*, hg. von Gerhard Allroggen, Berlin/Detmold, Version 3.1.0 (30. Juni 2017), <http://weber-gesamtausgabe.de> (letzter Aufruf 22.9.2017); die Musikwissenschaft hat die *Digitale Edition* allemal intensiv vorangetrieben, siehe etwa Peter Stadler/Joachim Veit (Hg.), *Digitale Edition zwischen Experiment und Standardisierung. Musik – Text – Codierung*, Tübingen 2009.

⁹⁷<http://www.tei-c.org/> (letzter Aufruf 22.9.2017) mit den aktuellen Guidelines P5.

⁹⁸Musil 1992 (Anm. 94), Musil 2009 (Anm. 94).

⁹⁹Vgl. Sahle (Anm. 19), Bd. 2, 157–165.

die Benutzeroberfläche, das *Interface*, ausrichtet.¹⁰⁰ Ob also in einigen Jahren die Edition noch so aufrufbar ist, wie sie ursprünglich ausgesehen hat, ist keineswegs sicher – und damit auch nicht, ob die Diskussionen, die anhand des Erscheinungsbilds einer *Digitalen Edition* heute geführt werden, in einigen Jahren noch nachvollzogen werden können. So wird sich das Problem der Langzeitarchivierung vielleicht noch als die letztlich entscheidende Frage für das Arbeitsgebiet der *Digitalen Historisch-kritischen Ausgabe* erweisen, eines Ausgabentypus, der von seinem wissenschaftlichen Anspruch her im Sinne der Grundlagenforschung auf lange Dauer – viele Jahrzehnte, wenn nicht Jahrhunderte – ausgerichtet ist. So dient die Weimarer *Goethe-Ausgabe* in ihren bis heute nicht ersetzten umfangreichen Teilen noch immer als wissenschaftlich verbindliche Referenz. Ihre 143 Bände haben vor 130 Jahren zu erscheinen begonnen und wurden zwischen 1887 und 1919 veröffentlicht.¹⁰¹ Wird entsprechend im Jahr 2147 eine *Digitale Historisch-kritische Ausgabe* aus dem Jahr 2017 noch nutzbar sein, ja überhaupt noch existieren?

Literatur

- Sämtliche digitalen Referenzen wurden letztmalig am 22.9.2017 eingesehen.
- Altenhöner, Reinhard, „Trau, schau, wem. Zur Authentizität und Langzeitverfügbarkeit digitaler Objekte“, in: *Bibliothek und Wissenschaft* 44 (2011), 153–170.
- Anz, Thomas (Hg.), *Handbuch Literaturwissenschaft. Gegenstände – Konzepte – Institutionen*, 3 Bde., Bd. 2: *Methoden und Theorien*, Stuttgart/Weimar 2007.
- Apollon, Daniel/Bélisle, Claire/Régnier, Philippe, „Introduction. As Texts Become Digital“, in: Dies. (Hg.), *Digital Critical Editions*, Urbana/Chicago/Springfield 2014, 1–32.
- Backmann, Reinhold, „Die Gestaltung des Apparates in den kritischen Ausgaben neuerer deutscher Dichter. (Mit besonderer Berücksichtigung der großen Grillparzer-Ausgabe der Stadt Wien)“, in: *Euphorion* 25 (1924), 629–662.
- Beckett, Samuel, *Samuel Beckett Digital Manuscript Project*, hg. von Dirk van Hulle/Mark Nixon/Vincent Neyt, Brüssel 2011 ff., <http://www.beckettarchive.org/>.
- Bentham, Jeremy, *Transcribe Bentham*, hg. von Philip Schofield/Melissa Terras/Tom Couch, London 2010 ff., <http://blogs.ucl.ac.uk/transcribe-bentham/>.
- Blake, William, *The William Blake Archive*, hg. von Morris Eaves/Robert N. Essick/Joseph Viscomi, <http://www.blakearchive.org/>.
- Boghardt, Martin, *Analytische Druckforschung. Ein methodischer Beitrag zu Buchkunde und Textkritik*, Hamburg 1977.
- Boghardt, Martin, *Archäologie des gedruckten Buches*, hg. von Paul Needham/Julie Boghardt, Wiesbaden 2008.
- Bohnenkamp, Anne/Henke, Silke/Jannidis, Fotis u. a., „Die digitale „Faust“-Edition. Zur neuen historisch-kritischen Ausgabe von Goethes Drama“, in: *Goethe-Jahrbuch* 133, 2016 [2017], 150–162.
- Büchner, Georg, *Woyzeck. Faksimileausgabe der Handschriften*, bearbeitet von Gerhard Schmid, Faksimile/Transkription, Wiesbaden 1981.

¹⁰⁰Siehe Reinhard Altenhöner, „Trau, schau, wem. Zur Authentizität und Langzeitverfügbarkeit digitaler Objekte“, in: *Bibliothek und Wissenschaft* 44 (2011), 153–170, hier: 167.

¹⁰¹Goethe (Anm. 50).

- Büchner, Georg, *Sämtliche Werke und Schriften. Historisch-kritische Ausgabe mit Quellendokumentation und Kommentar (Marburger Ausgabe)*, im Auftrag der Akademie der Wissenschaften und der Literatur, Mainz, hg. von Burghard Dedner, mitbegründet von Thomas Michael Mayer, Darmstadt 2000–2013.
- Büchner, Georg, *Georg-Büchner-Portal*, hg. von Burghard Dedner, Marburg, <http://buechnerportal.de/>.
- Burch, Thomas/Büdenbender, Stefan/Fink, Kristina u. a., „Text[ge]schichten. Herausforderungen textgenetischen Edierens bei Arthur Schnitzler“, in: Katharina Krüger/Elisabetta Mengaldo/Eckhard Schumacher (Hg.), *Textgenese und digitales Edieren. Wolfgang Koeppens ‚Jugend‘ im Kontext der Editionsphilologie*, Berlin/Boston 2016, 87–105.
- Burger, Hermann, *Lokalbericht. Digitale Edition*, hg. von Peter Dängeli/Magnus Wieland/Irmgard M. Wirtz u. a., Beta-Version, Bern 2016, <http://www.lokalbericht.ch/>.
- Dedner, Burghard, „Die Darstellung von Quellenabhängigkeiten anhand von Beispielen“, in: *editio* 11 (1997), 97–115.
- Dickinson, Emily, *Dickinson Electronic Archives*, hg. von Martha N. Smith/Marta Werner/Jessica Beard u. a., <http://www.emilydickinson.org/>.
- Escher, Alfred, *Digitale Briefedition*, hg. von Joseph Jung, Launch Juli 2015 (laufend aktualisiert), Zürich, <https://www.briefedition.alfred-escher.ch/>.
- Fanta, Walter, „Die Computer-Edition des Musil-Nachlasses. Baustein einer Epochendatenbank der Moderne“, in: *editio* 8 (1994), 127–157.
- Fontane, Theodor, *Notizbücher. Digitale genetisch-kritische und kommentierte Edition*, hg. von Gabriele Radecke, <https://fontane-nb.dariah.eu/index.html> (Version 0.1 vom 7. Dezember 2015, im Folgenden weiter aktualisiert).
- Gabler, Hans W., „Das wissenschaftliche Edieren als Funktion der Dokumente“, in: *Jahrbuch für Computerphilologie* 8/2006, <http://computerphilologie.digital-humanities.de/jg06/gabler.html>.
- Goethe, Johann W. von, *Goethes Iphigenie auf Tauris. In vierfacher Gestalt*, hg. von Jakob Baechtold, Freiburg i.Br./Tübingen 1883; zweite Ausgabe [als Titeldruck], Freiburg i.Br. 1888.
- Goethe, Johann W. von, *Goethes Werke*, hg. im Auftrage der Großherzogin Sophie von Sachsen, 143 Bde., Weimar 1887–1919.
- Goethe, Johann W. von, *Werke Goethes*, hg. von der Deutschen Akademie der Wissenschaften zu Berlin, Berlin 1952–1966, abgebrochen, danach weitere Bände als Einzelausgaben; Bd.: *Die Leiden des jungen Werthers*, 1: *Text. Erste und zweite Fassung*, Bearbeiter des Bandes: Erna Merker, Berlin 1954; Bd.: *Faust*, Ergänzungsband 3: *Urfaust – Faust. Ein Fragment – Faust. Der Tragödie Erster Theil (Paralleldruck)*, Bearbeiter des Bandes: Ernst Grumach, Inge Jensen, Berlin 1958.
- Goethe, Johann W. von, *Der junge Goethe in seiner Zeit. Texte und Kontexte*, 2 Bde. und CD-ROM, hg. von Karl Eibl/Fotis Jannidis/Marianne Willems, Frankfurt a. M./Leipzig 1998.
- Goethe, Johann W. von, *Proypläen. Goethes Biographica*, hg. von Bernhard Fischer/Klaus Manger/Ernst Osterkamp, Weimar/Frankfurt a. M. 2015 ff., <http://www.goethe-biographica.de/>.
- Goethe, Johann W. von, *Faust. Historisch-kritische Edition*, hg. von Anne Bohnenkamp/Silke Henke/Fotis Jannidis u. a., Beta-Version 3, Frankfurt a. M./Weimar/Würzburg 2016/17, <http://beta.faustedition.net/>.
- Gogh, Vincent van, *The Letters*, hg. von Leo Jansen/Hans Luijten/Nienke Bakker, Amsterdam/Den Haag 2009 (Version: April 2017), <http://vangoghletters.org>.
- Göttsche, Dirk, „Ausgabentypen und Ausgabenbenutzer“, in: Rüdiger Nutt-Kofoth/Bodo Plachta/H.T.M. van Vliet u. a. (Hg.), *Text und Edition. Positionen und Perspektiven*, Berlin 2000, 37–63.
- Groddeck, Wolfram/Sattler, D.E., „Frankfurter Hölderlin-Ausgabe. Vorläufiger Editionsbericht“, in: *Le pauvre Holterling. Blätter zur Frankfurter Hölderlin-Ausgabe* 2 (1977), 5–19.
- Handke, Peter, *Handkeonline*, hg. von Klaus Kastberger/Katharina Pektor/Christoph Kepplinger-Prinz, <http://handkeonline.onb.ac.at/>.

- Haugen, Odd E./Apollon, Daniel, „The Digital Turn in Textual Scholarship. Historical and Typological Perspectives“, in: Daniel Apollon/Claire Belisle/Philippe Regnier (Hg.), *Digital Critical Editions*, Urbana/Chicago/Springfield 2014, 35–57.
- Heine, Heinrich, *Säkularausgabe. Werke, Briefwechsel, Lebenszeugnisse*, hg. von den Nationalen Forschungs- und Gedenkstätten der klassischen deutschen Literatur in Weimar [seit 1993: Stiftung Weimarer Klassik, seit 2008: Klassik Stiftung Weimar] und dem Centre National de la Recherche Scientifique in Paris, Berlin/Paris 1970f ff.
- Heine, Heinrich, *Historisch-kritische Gesamtausgabe der Werke*, hg. von Manfred Windfuhr im Auftrag der Landeshauptstadt Düsseldorf, 16 Bde. in 23, Hamburg 1973–1997.
- Heine, Heinrich, *Heinrich-Heine-Portal*, hg. von Thomas Burch/Joseph A. Kruse/Bernd Füllner u. a. 2002–2009, <http://www.heine-portal.de/>.
- Heym, Georg, *Gedichte 1910–1912. Historisch-kritische Ausgabe aller Texte in genetischer Darstellung*, hg. von Günter Dammann/Gunter Martens/Karl L. Schneider u. a., 2 Bde., Tübingen 1993.
- Hölderlin, Friedrich, *Sämtliche Werke. ‚Frankfurter Ausgabe‘. Historisch-kritische Ausgabe*, hg. von D.E. Sattler, Frankfurt a. M. (seit 1985: Basel/Frankfurt a. M.) 1975–2008.
- Höpker-Herberg, Elisabeth/Zeller, Hans, „Der Kommentar, ein integraler Bestandteil der historisch-kritischen Ausgabe?“, in: *editio* 7 (1993), 51–61.
- Hulle, Dirk Van, „Modelling a Digital Scholarly Edition for Genetic Criticism. A Rapprochement“, in: *Variants* 12/13 (2016 [2017]), 34–56, <http://variants.revues.org/293>.
- Hurlbusch, Klaus, *Klopstock, Hamann und Herder als Wegbereiter autorzentrischen Schreibens. Ein philologischer Beitrag zur Charakterisierung der literarischen Moderne*, Tübingen 2001.
- Jannidis, Fotis, „Elektronische Edition“, in: Rüdiger Nutt-Kofoth/Bodo Plachta (Hg.), *Editionen zu deutschsprachigen Autoren als Spiegel der Editions-geschichte*, Tübingen 2005, 457–470.
- Jannidis, Fotis, „Computerphilologie“, in: Thomas Anz (Hg.), *Handbuch Literaturwissenschaft. Gegenstände – Konzepte – Institutionen*, 3 Bde., Bd. 2: *Methoden und Theorien*, Stuttgart/Weimar 2007, 27–40.
- Jannidis, Fotis, „Digitale Editionen“, in: *Mitteilungen des Deutschen Germanistenverbandes* 55/3 (2008), 317–332.
- Jannidis, Fotis/Kohle, Hubertus/Rehbein, Malte (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart 2017.
- Kafka, Franz, *Historisch-Kritische Ausgabe sämtlicher Handschriften, Drucke und Typoskripte*, hg. von Roland Reuß/Peter Staengle, Basel/Frankfurt a. M. 1995 ff.
- Kamzelak, Roland (Hg.), *Computergestützte Text-Edition*, Tübingen 1999.
- Kamzelak, Roland S., „Edition und EDV. Neue Editionspraxis durch Hypertext-Editionen“, in: Rüdiger Nutt-Kofoth/Bodo Plachta/H.T.M. van Vliet u. a. (Hg.), *Text und Edition. Positionen und Perspektiven*, Berlin 2000, 65–80.
- Kanzog, Klaus, *Prolegomena zu einer historisch-kritischen Ausgabe der Werke Heinrich von Kleists. Theorie und Praxis einer modernen Klassiker-Edition*, München 1970.
- Kastberger, Klaus/Kepplinger-Prinz, Christoph, „Handkeonline. Eine Forschungsplattform zu Peter Handke“, in: *editio* 27 (2013), 205–215.
- Keller, Gottfried, *Sämtliche Werke. Historisch-Kritische Ausgabe*, hg. unter der Leitung von Walter Morgenthaler im Auftrag der Stiftung Historisch-Kritische Gottfried Keller-Ausgabe, Basel/Frankfurt a. M./Zürich 1996–2013, digitaler Teil auf beigegebenen CD-ROMs und <https://www.ehka.ch/ehkka/>.
- Kleist, Heinrich v., *Sämtliche Werke*, [Berliner, seit 1992:] *Brandenburger Ausgabe*, hg. von Roland Reuß/Peter Staengle, Basel/Frankfurt a. M. 1988–2010.
- Koepfen, Wolfgang, *Jugend. Textgenetische Edition*, hg. von Katharina Krüger/Elisabetta Mengaldo/Eckhard Schumacher, 2017, <http://www.koepfen-jugend.de>.
- Kondrup, Johnny, „Text und Werk – zwei Begriffe auf dem Prüfstand“, in: *editio* 27 (2013), 1–14.
- Kraft, Herbert, „Die Edition fragmentarischer Werke“, in: *LiLi. Zeitschrift für Literaturwissenschaft und Linguistik* 5/19–20 (1975), 142–146.

- Kraft, Herbert, *Editionsphilologie*, mit Beiträgen von Jürgen Gregolin u. a., Darmstadt 1990; Zweite, neubearbeitete und erw. Aufl. mit Beiträgen von Diana Schilling/Gert Vonhoff, Frankfurt a. M. u. a. 2001.
- Lukas, Wolfgang, „Was ist das Digitalisierungsinteresse der geistes- und kulturwissenschaftlichen Forschung?“, in: LVR-Archivberatungs- und Fortbildungszentrum (Hg.), *Digital und analog. Die beiden Archivwelten*. 46. Rheinischer Archivtag. Ratingen 21.–22. Juni 2012. Beiträge, Bonn 2013, 32–47.
- Martens, Gunter, „Textdynamik und Edition. Überlegungen zur Bedeutung und Darstellung variierender Textstufen“, in: Gunter Martens/Hans Zeller (Hg.), *Texte und Varianten. Probleme ihrer Edition und Interpretation*, München 1971, 165–201.
- Martens, Gunter, „Text“, in: Klaus Kanzog/Achim Masser (Hg.), *Reallexikon der deutschen Literaturgeschichte*, Bd. 4: *Sl–Z*, Berlin/New York 2¹⁹⁸⁴, 403–417.
- Martens, Gunter, „Was ist ein Text? Ansätze zur Bestimmung eines Leitbegriffs der Textphilologie“, in: *Poetica* 21, 1989, 1–25.
- Martens, Gunter, „Was ist – aus editorischer Sicht – ein Text? Überlegungen zur Bestimmung eines Zentralbegriffs der Editionsphilologie“, in: Siegfried Scheibe/Christel Laufer (Hg.), *Zu Werk und Text. Beiträge zur Textologie*, Berlin 1991, 135–156.
- Martens, Gunter, „Das Werk als Grenze. Ein Versuch zur terminologischen Bestimmung eines editorischen Begriffs“, in: *editio* 18, 2004, 175–186.
- Mittler, Elmar/Rehbein, Malte, „Edition und Forschungsbibliothek. Chancen und Herausforderungen einer traditionsreichen Partnerschaft im digitalen Zeitalter“, in: *Bibliothek und Wissenschaft* 44 (2011), 9–21.
- Musil, Robert, *Der literarische Nachlaß*, hg. von Friedbert Aspetsberger/Karl Eibl/Adolf Frisé, Reinbek 1992 [CD-ROM].
- Musil, Robert, *Klagenfurter Ausgabe. Kommentierte digitale Edition sämtlicher Werke, Briefe und nachgelassener Schriften. Mit Transkriptionen und Faksimiles aller Handschriften*, hg. von Walter Fanta/Klaus Amann/Karl Corino, Klagenfurt, DVD-Version 2009.
- Neuber, Frederike, „Typografie und Varianz in Stefan Georges Werk. Konzeptionelle Überlegungen zu einer ‚typografiekritischen‘ Edition“, in: *editio* 31 (2017), 205–232.
- Nietzsche, Friedrich, *Werke. Kritische Gesamtausgabe*, begründet von Giorgio Colli/Mazzino Montinari, weitergeführt von Wolfgang Müller-Lauter/Karl Pestalozzi, Berlin u. a. 1967 ff.
- Nietzsche, Friedrich, *Digitale Faksimile-Gesamtausgabe. Nach den Originalmanuskripten und Originaldrucken der Bestände der Klassik Stiftung Weimar*, hg. von Paolo D’Iorio, Nietzsche Source, Paris 2009 ff., <http://www.nietzschesource.org/DFGA/>.
- Nietzsche, Friedrich, *Digitale Kritische Gesamtausgabe. Werke und Briefe, Nietzsche Source*, hg. von Paolo D’Iorio, Paris 2009 ff., Berlin/New York, 1967 ff., <http://www.nietzschesource.org/#eKGWB>.
- Nühlen, Kathrin/Wolf, Jonas, „Ansichtssache. Möglichkeiten der Darstellung und Interpretation textgenetischer Varianz am Beispiel der historisch-kritischen Edition ‚Arthur Schnitzler digital‘“, in: Wernfried Hofmeister/Andrea Hofmeister-Winter (Hg.), *Textrevisionen. Beiträge der Internationalen Fachtagung der Arbeitsgemeinschaft für germanistische Edition, Graz, 17.–20. Februar 2016*, Berlin/Boston 2017, 185–198.
- Nutt-Kofoth, Rüdiger (Hg.), *Dokumente zur Geschichte der neugermanistischen Edition*, Tübingen 2005.
- Nutt-Kofoth, Rüdiger, „Editionsphilologie“, in: Thomas Anz (Hg.), *Handbuch Literaturwissenschaft. Gegenstände – Konzepte – Institutionen*, 3 Bde., Bd. 2: *Methoden und Theorien*, Stuttgart/Weimar 2007, 1–27.
- Nutt-Kofoth, Rüdiger, „Editorische Axiome“, in: *editio* 26 (2012), 59–71.
- Nutt-Kofoth, Rüdiger, „Sichten – Perspektiven auf Text“, in: Anne Bohnenkamp (Hg.), *Medienwandel / Medienwechsel in der Editionswissenschaft*, Berlin/Boston 2013, 19–29.
- Nutt-Kofoth, Rüdiger, „Konzepte der Fragmentedition und ihre Probleme“, in: Matthias Berning/Stephanie Jordans/Hans Kruschwitz (Hg.), *Fragment und Gesamtwerk. Relationsbestimmungen in Edition und Interpretation*, Kassel 2015, 13–27.

- Nutt-Kofoth, Rüdiger, „Wie werden neugermanistische (historisch-)kritische Editionen für die literaturwissenschaftliche Interpretation genutzt? Versuch einer Annäherung aufgrund einer Auswertung germanistischer Periodika“, in: Thomas Bein (Hg.), *Vom Nutzen der Editionen. Zur Bedeutung moderner Editorik für die Erforschung von Literatur- und Kulturgeschichte*, Berlin/Boston 2015, 233–245.
- Nutt-Kofoth, Rüdiger, „Briefe herausgeben: Digitale Plattformen für Editionswissenschaftler und die Grundfragen der Briefedition“, in: Kristina Richts/Peter Stadler (Hg.), *„Ei, dem alten Herrn zoll' ich Achtung gern“: Festschrift für Joachim Veit zum 60. Geburtstag*, München 2016, 575–586.
- Nutt-Kofoth, Rüdiger, „Typographie als Informationssystem. Zum Layout der neugermanistischen Edition“, in: Rainer Falk/Thomas Rahn (Hg.), *Typographie & Literatur*, Frankfurt a. M./Basel 2016, 349–368.
- Nutt-Kofoth, Rüdiger, „Autorschaft, Werk, Medialität. Editionstheoretische Annäherungen an pluriautorschaftliche und plurimediale Werkkomplexe – mit einem germanistischen Blick auf das Phänomen Oper/Libretto“, in: Thomas Betzwieser/Norbert Dubowy/Andreas Münzmay u. a. (Hg.), *Perspektiven der Edition musikdramatischer Texte*, Berlin/Boston 2017, 25–38.
- Nutt-Kofoth, Rüdiger, „Textgenese analog und digital: Ziele, Standards, Probleme“, in: Anke Bosse/Walter Fanta (Hg.), *Textgenese im digitalen Medium*, Berlin/Boston 2019, 1–19.
- Pierazzo, Elena, *Digital Scholarly Editing. Theories, Models and Methods*, Farnham, Surrey/Burlington, VT, 2015.
- Plachta, Bodo, „Edition und Bibliothek“, in: *Bibliothek und Wissenschaft* 44 (2011), 23–36.
- Plachta, Bodo, *Editionswissenschaft. Eine Einführung in Methode und Praxis der Edition neuerer Texte*, 3., ergänzte und aktualisierte Aufl., Stuttgart 2013.
- Price, Kenneth M., „Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?“, in: *Digital Humanities Quarterly* 3/3 (2009), <http://www.digitalhumanities.org/dhq/vol/3/3/000053/000053.html>.
- Radecke, Gabriele/Neuroth, Heike/de la Iglesia, Martin u. a., „Vom Nutzen digitaler Editionen. Die Genetisch-kritische Hybrid-Edition von Theodor Fontanes Notizbüchern erstellt mit der Virtuellen Forschungsumgebung TextGrid“, in: Thomas Bein (Hg.), *Vom Nutzen der Editionen. Zur Bedeutung moderner Editorik für die Erforschung von Literatur- und Kulturgeschichte*, Berlin/Boston 2015, 277–295.
- Régnier, Philippe, „Ongoing Challenges for Digital Critical Editions“, in: Ders./Daniel Apollon/Claire Belisle (Hg.), *Digital Critical Editions*, Urbana/Chicago/Springfield 2014, 58–80.
- Reuß, Roland, „Text, Entwurf, Werk“, in: *Text. Kritische Beiträge* 10 (2005): *Text · Werk*, 1–12.
- Robinson, Peter, „Towards a Theory of Digital Editions“, in: *Variants* 10 (2013), 105–131.
- Robinson, Peter, *Why digital humanists should get out of textual scholarship / What digital humanists don't know about scholarly editing; what scholarly editors don't know about the digital world*, http://www.academia.edu/4124828/SDSE_2013_why_digital_humanists_should_get_out_of_textual_scholarship.
- Rossetti, Dante G., *Rossetti Archive*, hg. von Jerome J. McGann, <http://www.rossettiarchive.org/>.
- Sahle, Patrick, „Digitales Archiv – Digitale Edition. Anmerkungen zur Begriffsklärung“, in: Michael Stolz/Lucas Marco Gisi/Jan Loop (Hg.), *Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien. Eine Standortbestimmung*, Zürich 2007, 64–84.
- Sahle, Patrick, *A Catalog of Digital Scholarly Editions*, v 3.0, snapshot 2008ff. (letzte Änderung 22. März 2017), <http://www.digitale-edition.de/>.
- Sahle, Patrick, „Zwischen Mediengebundenheit und Transmedialisierung. Anmerkungen zum Verhältnis von Edition und Medien“, in: *editio* 24 (2010), 23–36.
- Sahle, Patrick, *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, 3 Bde., Bd. 1: *Das typographische Erbe*, Bd. 2: *Befunde, Theorie und Methodik*, Bd. 3: *Textbegriffe und Recodierung*, Norderstedt 2013.
- Sahle, Patrick, „What is a Scholarly Digital Edition?“, in: Matthew James Driscoll/Elena Pierazzo (Hg.), *Digital Scholarly Editing. Theories and Practices*, Cambridge, UK, 2016, 19–39.

- Sahle, Patrick, „Digitale Edition“, in: Fotis Jannidis/Hubertus Kohle/Malte Rehbein (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart 2017, 234–249.
- Sauer, Alfred/Seuffert, Bernhard, *Briefwechsel*, hg. von Bernhard Fetz/Hans H. Müller/Desiree Hebenstreit u. a., Wien/Hamburg 2016, <http://sauer-seuffert.onb.ac.at/>.
- Scheibe, Siegfried, „Zum editorischen Problem des Textes“, in: *Zeitschrift für deutsche Philologie* 101 (1982), Sonderheft: Norbert Oellers/Hartmut Steinecke (Hg.), *Probleme neugermanistischer Edition*, 12–29.
- Scheibe, Siegfried, „Editorische Grundmodelle“, in: Siegfried Scheibe/Christel Laufer (Hg.), *Zu Werk und Text. Beiträge zur Textologie*, Berlin 1991, 23–48.
- Schiller, Friedrich, *Schillers Werke. Nationalausgabe*, 1940 begründet von Julius Petersen, fortgeführt von Lieselotte Blumenthal, Benno von Wiese, Siegfried Seidel, hg. von Norbert Oellers im Auftrag der Stiftung Weimarer Klassik und des Schiller-Nationalmuseums in Marbach, Bd. 5.: Neue Ausgabe: *Kabale und Liebe, Semele, Der versöhnte Menschenfeind, Körners Vormittag*, hg. von Herbert Kraft/Claudia Pilling/Gert Vonhoff, Weimar 2000.
- Schlegel, August W., *Digitale Edition der Korrespondenz August Wilhelm Schlegels*, Projektleitung: Jochen Strobel u. a., <http://august-wilhelm-schlegel.de/briefedigital/>, Beta-Version, Stand der Daten: 4.7.2017.
- Schönert, Jörg, „Literaturgeschichtsschreibung“, in: Thomas Anz (Hg.), *Handbuch Literaturwissenschaft. Gegenstände – Konzepte – Institutionen*, 3 Bde., Bd. 2: *Methoden und Theorien*, Stuttgart/Weimar 2007, 267–284.
- Shakespeare, William, *Shakespeare Electronic Archive*, hg. von Peter S. Donaldson/Larry Friedlander/Janet H. Murray u. a., Cambridge, MA, <http://shea.mit.edu/>.
- Shelley, Percy B./Wollstonecraft Shelley, Mary/Godwin, William/Wollstonecraft, Mary, *The Shelley-Godwin Archive*, hg. von Neil Fraistat/Elizabeth Denlinger/Raffaele Vigiante u. a., <http://shelleygodwinarchive.org/>.
- Shillingsburg, Peter, „Principles for Electronic Archives, Scholarly Editions, and Tutorials“, in: Richard J. Finneran (Hg.), *The Literary Text in the Digital Age*, Ann Arbor 1996, 23–35.
- Shillingsburg, Peter L., *From Gutenberg to Google. Electronic Representations of Literary Texts*, Cambridge, UK/New York u. a. 2006.
- Shillingsburg, Peter, „Development Principles for Virtual Archives and Editions“, in: *Variants* 11 (2014), 11–28.
- Stäcker, Thomas, „Creating the Knowledge Site – elektronische Editionen als Aufgabe einer Forschungsbibliothek“, in: *Bibliothek und Wissenschaft* 44 (2011), 107–126.
- Stadler, Peter/Veit, Joachim (Hg.), *Digitale Edition zwischen Experiment und Standardisierung. Musik – Text – Codierung*, Tübingen 2009.
- Strohschneider Peter/Vollhardt, Friedrich, „Interpretation. Eine Einleitung in den Thementeil dieses Heftes“, in: *Mitteilungen des Deutschen Germanistenverbandes* 49/2 (2002): *Interpretation*, 98–102.
- Trakl, Georg, *Sämtliche Werke und Briefwechsel. Innsbrucker Ausgabe. Historisch-kritische Ausgabe mit Faksimiles der handschriftlichen Texte Trakls*, hg. von Eberhard Saueremann/Hermann Zwerschina, Basel/Frankfurt a. M. 1995–2014.
- Vanhoutte, Edward, „Where is the editor? Resistance in the creation of an electronic critical edition“, in: *Human IT* 1 (1999), 197–214, <http://etjanst.hb.se/bhs/ith/1-99/ev.htm>.
- Weber, Carl Maria von, *Gesamtausgabe. Digitale Edition*, hg. von Gerhard Allroggen, Berlin/Detmold, Version 3.1.0 (30. Juni 2017), <http://weber-gesamtausgabe.de>.
- Winko, Simone, „Textbewertung“, in: Thomas Anz (Hg.), *Handbuch Literaturwissenschaft. Gegenstände – Konzepte – Institutionen*, 3 Bde., Bd. 2: *Methoden und Theorien*, Stuttgart/Weimar 2007, 233–266.
- Zeller, Hans, „Edition und Interpretation. Antrittsvorlesung“, in: *zürcher student* 43/7 (Januar 1966), 15 u. 19.
- Hans Zeller, „Befund und Deutung. Interpretation und Dokumentation als Ziel und Methode der Edition“, in: Gunter Martens/Hans Zeller (Hg.), *Texte und Varianten. Probleme ihrer Edition und Interpretation*, München 1971, 45–89.

- Zeller, Hans, „Für eine historische Edition. Zu Textkonstitution und Kommentar“, in: Georg Stötzel (Hg.), *Germanistik – Forschungsstand und Perspektiven. Vorträge des Deutschen Germanistentages 1984*, 2. Teil: *Ältere Deutsche Literatur, Neuere Deutsche Literatur*, Berlin/New York 1985, 305–323.
- Zeller, Hans, „Die Faksimile-Ausgabe als Grundlagenedition für Philologie und Textgenetik. Ein Vorschlag“, in: Hans Zeller/Gunter Martens (Hg.), *Textgenetische Edition*, Tübingen 1998, 80–100.

Online-Ressourcen

Nietzsche Source, <http://www.nietzschesource.org/>.

Text Encoding Initiative, <http://www.tei-c.org/>.

Vernetzte Korrespondenzen – Exilnetz33, <http://exilnetz33.de>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Diskussionsbericht Sektion II. Digitale Edition und Annotation

Luise Borek

Die sieben Beiträge der Sektion *Digitale Edition und Annotation* befassten sich mit dem Edieren von Texten im digitalen Raum, dem Prozess ihrer Aufbereitung und den dabei verwendeten Methoden. Die Reflexion dieser Praxis bezieht sich nicht nur auf einen medialen Transformationsprozess – sie erlaubt und erfordert es auch, grundlegende philologische Aktivitäten neu zu verhandeln.

Während die Edition gleichermaßen ein Forschungsergebnis und die Grundlage für weitere Forschung darstellt, sind Annotationen Teil eines aktiven Leseprozesses und Elemente eines analysierenden Zugriffs. Beide stehen somit für philologische Aushandlungsprozesse zwischen Deskription und Interpretation.

Im Inkunabelzeitalter der *Digitalen Edition* stellt es eine besondere Herausforderung dar, den andauernden Entwicklungsprozess ins Digitale parallel zu reflektieren, zu theorisieren und zu standardisieren. Dabei geht es einerseits um das Verhältnis analoger und digitaler Formate und die mit ihm einhergehenden Konsequenzen für LeserInnen und HerausgeberInnen. Welches sind die wesentlichen und identitätsstiftenden Elemente der Edition, und was leistet ihre digitale Umsetzung? Andererseits geht es in der Diskussion immer wieder um das Explizieren von Implizitem und damit um das Verhältnis von Modellierung, Abstraktion und Hermeneutik. In diesem Zusammenhang wurde auch auf die Problematik des weit verbreiteten Begriffs des *Close Reading* hingewiesen, der gemeinhin als Gegensatz zu maschinellen Verfahren verstanden wird – und keinesfalls ein komplementärer Begriff zu Morettis *Distant Reading* ist, wie es die Begriffe suggerieren.

Der Beitrag von Matthias Bauer, Gabriel Viehhauser und Angelika Zirker skizzierte den Kommentar als Bindeglied zwischen objektiver Primärtextanalyse

L. Borek (✉)
TU Darmstadt, Darmstadt, Deutschland
E-Mail: luise.borek@tu-darmstadt.de

und Interpretation. In der Diskussion wurde an die lange Tradition philologischer Kommentierung erinnert, die zwar bislang nicht spezifisch digital praktiziert worden sei, aber eine bestehende Heuristik entwickelt habe, die auch im digitalen Format Bestand haben könne. Die Schwierigkeit des Transfers bestehe jedoch darin, dass sich die Tradition auf die Praxis beschränke und nur wenig Theorie liefere. Ein terminologisches Resultat dieses Umstands sei die Undifferenziertheit des Kommentarbegriffs, die ebenso auf die ‚Annotation‘ zutreffe. Die Beitragenden verstehen den Kommentar als erklärende Annotation und somit als einen Subtyp von Annotationen.

Ausgiebig diskutiert wurde das Spannungsfeld von Abgeschlossenheit und Offenheit von Kommentaren. Bereits der analogen Textsorte sei eine gewisse Offenheit inhärent, weil sich in ihr nicht ausschließlich auf das Objektive beschränkt werden könne und weil sie abhängig sei von Verstehensbedingungen, die sich fortlaufend aktualisierten. Es bestand Konsens darüber, dass der digitale Kommentar ein Potenzial zur weiteren Öffnung biete, das weit über wegfallende räumliche Beschränkungen hinausgehe. Ein praktisches Beispiel, dessen Verwirklichung auch im Digitalen noch ungelöst sei, bestehe in der Verankerung von Annotationen, die sich nicht auf spezifische Textstellen bezögen, sondern erst in einem größeren Kontext Relevanz erhielten oder sich aus einer Summe von Annotationen ergäben. Dies illustrierte zudem die Inklusion des hermeneutischen Prozesses in die Praxis des Kommentierens. Auch bei dem Verhältnis zwischen Kommentar und Interpretation müsse daher ein Mittelweg gefunden werden, der das öffnende Potenzial des Digitalen hin zu Multiperspektivität und dem Aufbrechen des Textbegriffs fördere, ohne den objektiven Gehalt des Kommentars zu unterwandern.

Weitere Diskussionsbeiträge widmeten sich dem Bereich der Systematisierung, dessen Wichtigkeit besonders hervorgehoben wurde. Das von der Autorin eingebrachte *TEASys-Modell* ermöglicht eine Kategorisierung der vorgenommenen Annotationen und setzt somit an dem Desiderat der Präzisierung von Annotationstypen und der Transparenz des Prozesses an. Auf die Frage nach der Bewährung in der Praxis wurde auf einen begleitenden *Living Style Guide* hingewiesen, der im Gegensatz zu weithin verbreiteten *Inter-Annotator Agreements* erforderliche Aushandlungsprozesse ermöglicht. Insbesondere im Umfeld der Lehre habe sich das kategorisierende Kommentieren als gutes reflexives Verfahren und produktive Erschließungsmethode erwiesen. Ausgehend von der Frage nach einer Systematisierung von Rollen, die im Beitrag nicht angesprochen wurde, aber spätestens bei einer Öffnung der Annotationsumgebung an Relevanz gewinnen werde, sei zu diskutieren, inwieweit es sich beim Digitalen Kommentar noch immer um eine Textsorte handle, die sehr stark an Experten gebunden sei.

Manuel Braun, Sonja Glauch und Florian Kragl untersuchten in ihrer Vorlage insbesondere jene Eigenschaften der *Digitalen Edition*, die über analoge Editionen hinausgehen. Sie gingen davon aus, dass Letztere inzwischen immer auch digital (angelegt) seien. Die vorgenommene Standortbestimmung verfolgte daher keineswegs das Ziel, das eine Format gegen das andere Format auszuspielen.

Ebenfalls kritisch hinterfragt wurde, inwiefern die Möglichkeit zur Nutzung quantitativer Verfahren tatsächlich zu neuen Qualitäten der *Digitalen Edition* beitragen könne, zumal sie große Herausforderungen an eine transparente und übersichtliche Darstellung mit sich bringe und zu keinen verlässlichen Ergebnissen führe. Während sich die Beitragenden dieser Herausforderungen bewusst waren, verwiesen sie auf das Umschlagen des quantitativen Arguments in ein qualitatives, sofern Zugänge geschaffen würden, die bereitgestellte Daten für individuelle Forschungsanfragen nutzbar machen würden. Das implizite Abwägen dieser Verfahrenskombinationen unterstreiche, dass sich die *Digitale Edition* in einem Grenzbereich zwischen Archiv und Edition bewege. Eine derartige Edition sei eine Analyseplattform, für die auch Maschinenlesbarkeit eine Rolle spiele.

Wie im Beitrag angeregt, wurde in diesem Zusammenhang diskutiert, inwiefern es sich bei der *Digitalen Edition* um ein revolutionäres Format handele. Ohne an dieser Terminologie festzuhalten, wurde übereinstimmend konstatiert, dass der Medienwandel hier zumindest zu neuen Qualitäten beitrage, zu denen neben der hinzugewonnenen Dynamik auch ein neues Rollenverständnis gehöre. So wurde in Replik auf den Beitrag auf eine Rollenverschiebung im Digitalen hingewiesen: Während die Herausgeberrolle vergleichsweise geschwächt zu sein scheine, indem dieser zunehmend organisierende Funktion zukomme, werde die Rolle des Rezipienten gestärkt. Im Beitrag war die Rede von BenutzerInnen als ‚Co-EditorInnen‘, was in der Diskussion hinterfragt wurde. Die Beitragenden erläuterten, dass sie es keineswegs so verständen, dass alles Machbare auch umgesetzt werden solle, sondern verwiesen darauf, dass entsprechende Zugänge geschaffen werden müssten, die unterschiedlichen Bedarfen gerecht würden. In der Diskussion wurde die Möglichkeit vielfältiger Angebote kritisch diskutiert, da hier nicht nur die Gefahr einer Überforderung von Nutzenden bestünde, sondern solche Surrogatlösungen auch deshalb sehr riskant seien, da das Angebot bequemer Alternativen ein Zurückfallen in analoge Muster bedeute, die es unbedingt zu vermeiden gelte. Als Beispiel wurde hier die hochwertige historisch-kritische Edition angeführt, die in der Regel um eine lesbarere und besser verkäufliche Studienausgabe ergänzt werde. Als Fazit der Diskussion der veränderten Rollen wurde festgehalten, dass die Digitale Edition LeserInnen stärker in die Pflicht nehme, als es bei ihrem gedruckten Pendant bislang der Fall gewesen sei. Es handele sich hier um eine Auslagerung, die zum gegenwärtigen Zeitpunkt als radikal empfunden werden könne.

Bei aller Einigkeit über den tatsächlichen und potenziellen Mehrwert *Digitaler Editionen* stellte sich die Frage nach dem Stellenwert der Textkritik, die hier in neuartiger Weise praktiziert werden könne, jedoch aus bislang nicht erklärlichen Gründen eher marginalisiert worden seien. In einigen Bereichen, wie beispielsweise den Aushandlungsprozessen zu Leithandschriften in der Mediävistik, sei die Vernachlässigung der Textkritik nachvollziehbar, da sich hier – etwa in Form von synoptischen Darstellungen – Möglichkeiten bieten würden, die Vielfalt der Überlieferung abzubilden. Darüber hinaus müsse die Entwicklung des laufenden Prozesses abgewartet werden, in dem eine Umkehrung des identifizierten Trends jederzeit eintreten könne.

Mit der Frage nach den Kosten für *Digitale Editionen* und deren Langzeitarchivierung widmete sich die Diskussion einem weiteren Bereich, der einige offene Fragen bereithalte, die es erst noch zu verhandeln gelte. Während Langzeitarchivierung nichts Neues sei, der Weg hier also als vergleichsweise geebnet angesehen werden könne und EditorInnen mit XML/TEI eine plausible Nachhaltigkeit schafften, wären aus der Perspektive der Förderer Mindeststandards für *Digitale Editionen* notwendig. In diesem Zusammenhang müsse zudem berücksichtigt werden, dass Schief lagen durch die Förderung entstehen könnten, etwa durch die Auswahl förderungswürdiger Editionsprojekte und den erhöhten Aufwand bei der Edition von älteren Texten. Kostenreduzierendes Potenzial zeige sich derweil im Bereich der Toolentwicklung zur nationalen Infrastruktur der Edition. Auch die Nutzung gemeinsamer Oberflächen könne die Kosten weiter reduzieren.

Für allgemeine Verwunderung sorgte das Resümee des Beitrags, in dem die *Digitale Edition* als zusätzliches und eigengesetzliches Angebot angesehen wird, das gegenüber seinem gedruckten Gegenstück nur für bestimmte literaturwissenschaftliche Anliegen einen Mehrwert biete. Die Beitragenden erläuterten ihr Fazit, indem sie analoge und digitale Edition als konzeptuelle Größen auffassen, die sich eignen, um bestimmte Praktiken und Zielsetzungen des Edierens zu umschreiben. Dass ein konkretes digitales Angebot in der Praxis beides umfassen und kombinieren kann, steht dazu in keinem Widerspruch.

Gerrit Brüning stellte in seiner Vorlage Überlegungen zu neuen Szenarien der Nutzung wissenschaftlicher Editionen unter den Voraussetzungen des Digitalen an. Es gehe darum, den im editionswissenschaftlichen Diskurs als ‚Variantenfriedhof‘ kritisierten Apparat digital zugänglich zu machen und für Analysen zu nutzen. Die Art seiner Aufbereitung solle in der Editionsphilologie diskutiert und nicht der Informatik überlassen werden.

Es sei nach allgemeinem Verständnis nicht die Aufgabe von Editionen, die Analyse und ihr Ergebnis vorzugeben, sondern lediglich, die entsprechenden Daten für die Analyse bereitzustellen. Die Analyse der Daten könne bereits mit grundlegenden Operationen erreicht werden, die zeigten, welche Änderungen an einem Text stattgefunden haben. Flankierend forderte er Werkzeuge zur Klassifikation von Typen von Varianten, etwa die Orthographie oder Interpretationen betreffend, ein. Hier wurde in der Diskussion darauf hingewiesen, dass bereits kleine Änderungen an einem Text diesen frapperend beeinflussen könnten. Scheinbar banale Eingriffe wie beispielsweise das Ändern numerischer in ausgedruckte Zahlen könnten eine Erzählstimme grundlegend verändern. Daraus wurde gefolgert, dass es keine übergreifende Typologie von Änderungen geben könne und eher textspezifische Kategorien zu erwarten seien.

Aus mediävistischer Perspektive werde der ‚Variantenapparat als Signum der Wissenschaft‘ wahrgenommen, dessen Zweck es sei, zu erklären, was im (darüberstehenden) Text passiere, und zu belegen, dass bei der Edition sauber gearbeitet worden sei. Die Rekonstruierbarkeit einer Violdimensionalität der Überlieferung sei dabei nicht das ausgegebene Ziel. Die Komplexitätsgrade zwischen Text und Apparat würden zudem variieren und es gebe keine einheitliche Vorgehensweise der HerausgeberInnen von Editionen. Aufgrund einer fehlenden Systematik

seien LeserInnen den Apparaten ausgeliefert, was zur Folge habe, dass diese mit ihnen meist nicht viel anfangen könnten. Während der Autor des Beitrags diese Zustandsbeschreibung für Editionen aus der neueren Literaturwissenschaft bestätigte, ging er davon aus, dass eine maschinelle Auswertung der Variantenapparate möglich sei.

Als ein denkbarer Ansatz wurde das Anlegen einer Variantenhermeneutik vorgeschlagen, mittels derer bereits im Vorfeld Aussagen zur Komplexität der jeweiligen Variantenlage getätigt werden könnten. Auf diese Art könnten gezielt Ansichten oder Zugänge gewählt werden – etwa eine bestimmte Phase der Überarbeitung eines Textes im Gegensatz zu einer komplexen Textgenese. Der Autor sah in diesem Plädoyer für die einfachen Ansätze, zu denen er auch den Vergleich von Druckfassungen zählte, eine Chance, sich der Darstellung und Analyse komplexerer Sachverhalte anzunähern und diesen den Weg zu ebnen.

Eine grundsätzliche Frage, die die Diskussion der Varianten ans Licht brachte, betraf das Verständnis von Fassungen. Bei ihrer Bestimmung handele es sich um einen interpretativen Prozess, der auf textgeschichtlichem Vorverständnis beruhe. Der Autor der Vorlage ging davon aus, dass jede Variante über das Potenzial verfüge, eine Fassung zu konstruieren. In diesem Zusammenhang wird darauf hingewiesen, dass es sich bei der Grundsicht in der Regel nicht um eine Fassung handele, die jemals historisch existiert habe. Vor diesem Hintergrund stellt sich auch die Frage, ob nicht eigentlich *Relationen* von Fassungen zu klassifizieren seien und nicht (oder nicht nur) Typen von Varianten.

Im Anschluss unternahm die Diskussion einen Perspektivwechsel: Gehe man davon aus, dass die EditorInnen die Überlieferung am besten kennen, stelle sich die Frage, ob nicht die Editionsphilologie der Zukunft diese Kenntnisse bis zu einem gewissen Grad in die Edition einfließen lassen sollte – etwa durch die Beigabe von Bewertungs- und Interpretationswegen, die zu diesem Zweck transparent gestaltet werden müssten. Auch auf eine damit einhergehende Verlagerung in der Methodologie der Editorik wurde in diesem Zusammenhang hingewiesen. Diese betreffe eine Reihe von Fragen, etwa danach, was auf digitaler Ebene ediert werde oder welche Leistungen EditorInnen für einen ‚unvordenklichen‘ – von unterschiedlichen Interessen geleiteten – Abfrageprozess erbringen müssten, um LeserInnen Genüge zu tun. Es handele sich somit auch um einen Übersetzungsprozess hin zur Maschinenlesbarkeit, für den relevant sei, welche konkreten Anforderungen an die Daten gestellt würden und welche Art der Deskription der Texte hierfür erforderlich sei. Inwieweit die Edition bereits Vorleistungen zu Abfragemöglichkeiten erbringen solle, sei auch abhängig von dem Bedarf der Forschenden und davon, welche Operationen diese selbst übernehmen wollen und können. Ob beispielsweise eine Lemmatisierung oder ein *Part-of-Speech Tagging* standardmäßig beigegeben sein solle, sei zudem abhängig von der Art der jeweiligen Edition. Eine linguistische Edition verfolge eine andere Zielsetzung als ein Lesetext etc. Gerrit Brüning formulierte die Bereitstellung aller textgeschichtlich existenter Tokens als eine Minimalanforderung und merkte an, dass diese für eine Volltextsuche ohnehin benötigt würden. Zugleich wies er darauf hin, dass in Anbetracht der

Verzeichnungsweise vieler vorliegender Editionen für die Ermittlung der Tokens bereits ein Aufbereitungsprozess notwendig sei.

Schließlich erfolgte die Anregung, mittels computerlinguistischer Verfahren quantitative Auswertungsmöglichkeiten zu eruieren. Zu den spontanen Ideen in dieser Richtung gehörten das Generieren eines unüberwachten Clusterings aus den vorgenommenen Änderungen oder die Zuordnung zu bereits gegebenen Klassen.

Anders als im vorherigen Paper ging es in der Vorlage von Dirk von Hulle um die Varianten des Autors selbst. In seinem Beitrag sprach er sich für die Aufwertung und für die editorische Berücksichtigung von verlorenem Material aus. Dieses finde sich häufig in Form von Notizbüchern, die nicht nur materielle Überlieferungsträger seien, sondern auch Indizien für einen Schaffensprozess liefern und somit Aufschluss über die Poetik eines Autors geben könnten.

Die Frage nach der Kodierung von Notizen, auf die AutorInnen für ihre Texte zurückgegriffen haben, ist einfach zu beantworten. Zunächst stehe hier aber der Versuch im Vordergrund, die Notizbücher überhaupt in die Edition zu integrieren. Der Autor verweist zudem auf die Schwierigkeit, dass ein Notizbuch nicht *einem* Text zuzuordnen sei, da es sich dabei um Work in Progress handele.

Daran schloss sich die Frage an, wieweit die Vorstellung eines nicht zielgerichteten Prozesses (in Anlehnung an den von Ernst Haeckel geprägten Begriff *Dysteleological Approach*) trage, zumal die Entscheidung für eine Übernahme in einen Text ja als durchaus zielgerichtet aufgefasst werden könne. Der Autor identifizierte diese Nachfrage als den Kern einer andauernden Diskussion, wobei Teleologie bei der Erstellung von Editionen immer implizit sei (daher spreche man auch von ‚Projekten‘). Der teleologische Anspruch sei zudem verankert und werde von HerausgeberInnen und LeserInnen gleichermaßen erwartet und gefordert. Man habe es jedoch mit übergangslosen Arbeiten zu tun, bei denen AutorInnen alles Mögliche notierten, ohne zu wissen, ob und für welchen Text sie darauf zurückgreifen würden. Inwiefern dabei unterschieden werden könne zwischen verwendetem und lediglich Notiertem und ob letzteres einen Kontext generiere, der zum Auswahlprozess gehöre, könne nicht eindeutig geklärt werden, weshalb hier Vorsicht geboten sei. In jedem Fall erlaube es die digitale Umgebung, diese Vorgänge besser nachzuvollziehen, indem neben der Präsentation von Notizbüchern mittels Transklusion auch teleologische Ansichten generiert werden könnten.

Während es sich bei dem Beckett-Projekt um ein Editionsprojekt handele, das TEI-Daten umfasse, basierten die dem Beitrag zugrunde gelegten *Notebook Editions* auf einer Datenbank-Struktur, die es erlaube, Kategorien anzusteuern und neben der Perspektive auf die Notizbücher auch flankierendes Material, wie z. B. die Bibliothek des Autors, einzubeziehen. In diesem Zusammenhang regte die Diskussion auch zu einer Rekonzeptualisierung des Vorhabens an, indem die Flexibilität der Perspektive betont wurde. Anstatt eine fixe Perspektive anzubieten, könnten hier ganz im Sinne der genetischen Kritik diachrone Ansichten generiert werden, zwischen denen flexibel gewechselt werden könne. Auf die Frage nach der Verlässlichkeit der verknüpften Informationen in der Modellierung führte Van Hulle aus, dass beim Versuch der Rekonstruktion einer *Virtual Library* Kriterien benötigt würden, die Grade der *Certainty* und *Uncertainty* wiedergäben. Auch

Spekulationen von EditorInnen, die intertextuelle Referenzen bemerkten, sollten als solche gekennzeichnet werden können.

Mit der Frage nach Material, das nicht vom Autor selbst stamme, etwa Notizen, die LeserInnen im Text hinterlassen haben, wurde eine weitere Dimension ins Spiel gebracht, deren Auswertung Aufschlüsse zu einer historischen Leserschaft geben könne. Van Hulle gab jedoch zu bedenken, dass es sich hier um eine sehr hypothetische Frage handle, die aus der aktuellen Projektperspektive nicht zu beantworten sei. Ein interessantes Szenario könne in dem Zusammenkommen von Notizen und gleichzeitigen Exzerpten liegen. Hier würde aus einem spezifischen Kontext ersichtlich, was ein lesender Autor als eigenes ‚Take-Away‘ aus einem Kontext herauslöse. Dabei handle es sich um Aspekte, die die Editions-wissenschaft bislang ziemlich ausgespart habe.

Es wurde angemerkt, dass es sich aus mediävistischer Perspektive bei den Spuren, die frühere LeserInnen in den Texten hinterlassen hätten, um absolut Wesentliches handle. Im Gegensatz zu diesen selteneren historischen Lesezeugnissen habe man es nun mit einer überwältigenden Datenmenge zu tun. Daraus leite sich die Frage ab, wo hier die Unterschiede zu verorten seien. Dirk Van Hulle sah diese vor allem in der Fluidität der Daten, in denen Kommunikation zwischen Beteiligten stattfinde und Veränderungen und Streichungen vorgenommen würden. Ein weiterer Austausch in diese Richtung und zum historischen Wandel der Annotationskultur wäre sehr willkommen.

Als interessanter Aspekt wurden neben den Konsequenzen, die diese Art der Edition hat, vor allem auch die Konsequenzen für Leser- und Nutzerrollen – und mit ihnen für Bibliotheken und Archive – angesehen. Mit Blick auf die jüngere Textproduktion gelte es zudem, Veränderungen im Schaffensprozess von *born-digital*-AutorInnen zu untersuchen, bei denen beispielsweise auch digitale Annotationen und Zeugnisse an die Stelle von Notizbüchern träten. Hierin sah der Verfasser der Vorlage eine ganz neue Herausforderung für die veränderte Editorenrolle. Das Problem skalieri zudem insofern, als man voraussichtlich kaum noch von ‚Fassungen‘ sprechen könne und man es mit einem zunehmend fluiden Prozess zu tun habe, in dem auch besuchte Websites und dergleichen eine Rolle spielen könnten. Die Auffassung von ‚Versionen‘ müsse in diesem Kontext völlig neu gedacht werden und sei eine spannende Herausforderung.

Mit der *Collation Engine*, die bislang zur Erstellung von Apparaten eingesetzt wurde, verbinde er einerseits die Hoffnung, den Apparat als solchen zu stärken, und wünsche sich darüber hinaus ihre Verwendung durch die NutzerInnen, denen hierfür entsprechende Wege aufgezeigt werden müssten.

Der Beitrag von Jan Christoph Meister enthielt ein Plädoyer für eine generelle Ausweitung des Funktionsumfangs von Annotationen. Die „selbstverordnete hermeneutische Abstinenz der *Digital Humanities*“ trage dazu bei, dass traditionelle und *Digitale Literaturwissenschaft* immer weiter auseinanderdrifteten. Dabei bestehe neben deklarativen Annotationspraktiken auch hermeneutisches Potenzial in Markup. Interpretation müsse als reflexive Praxis verstanden werden und Annotieren als Aushandlungsprozess des Verstehens. Mittels Annotation könnten Lesarten kommuniziert und offenere Interpretationen

aufgezeichnet werden. Vor diesem Hintergrund behandelte der Beitrag diskursives Annotieren in Forschungsumgebungen und formulierte Desiderate für den Umgang mit unterschiedlichen Modellierungsszenarien.

Die Diskussion der Vorlage erweckte den Wunsch nach einem Beispiel von hermeneutischem Markup. Momentan könne diesem nicht entsprochen werden, da hierfür zunächst *Hermeneutic Ground Challenges* definiert werden müssten, die ein Set von Problemen darstellten, um Operationalisierbarkeit systematisch zu testen. Der Autor der Vorlage erläutert zudem, dass das Annotieren hermeneutisch werde, sobald es methodenreflexiv sei und zu Problemen führe. Dies äußere sich beispielsweise in dem Bedarf, Entscheidungen zu plausibilisieren. Das Kollaborative könne hier einen wichtigen Beitrag leisten, indem unterschiedliche Bewertungen ausgetauscht und diskursiv ausgehandelt würden. Das etablierte *Inter-Annotator Agreement* sei diesbezüglich sehr dysfunktional, die Reflexion solle vielmehr in den Annotationsvorgang aufgenommen werden, um hier Diskursivität herzustellen.

Eine ausgedehnte Diskussion entwickelte sich um die Kontinuitätsvorstellung von Markup, das diskursive Elemente integriere, und um die Frage, ob hier nicht zwischen verschiedenen Markupformen zu unterscheiden wäre, damit eine zyklische Entwicklung von Statischem getrennt werden könne. Ein strategisches Argument für die Integration diskursiver Elemente liege in der Vermittlung. Annotieren habe immer etwas Stetiges, dennoch sei der Übergang zu hermeneutischen Fragen fließend. Komplexe Phänomene würden jedoch nur dann erreicht, wenn sie nicht ins angelegte Modell passen und somit nicht mehr leicht zu annotieren seien. Bei der intuitiven Annotation handele es sich um eine ‚Durchgangsphase‘, die transzendiert werden müsse. Es gehe folglich um eine Herausforderung des Konzepts. Ausgehend von einem klassenlosen Ursprung solle über die Begegnung und Auseinandersetzung mit einem geregelten Kontext aufgezeigt werden, dass es Kategorien geben *könne* – diese könnten jederzeit erweiterbar sein und auch meta-annotiert werden. Bei der Implementierung gebe es klare technische Grenzen. Diskontinuierliche Annotationen stellten hier kein triviales Problem dar, eine ausgehebelte Ebenenhistorie sei schwierig zu öffnen für analytische Verfahren. Für das Nachdenken und im theoretischen Metadiskurs sei das kategoriale System notwendig und solle als wichtiges Reflexionsinstrument nicht aufgegeben werden. Meister wies in diesem Zusammenhang auch darauf hin, dass man nicht darauf beharren solle, dass etwas nicht modellierbar sei – in diesem Fall würde jeglicher Dialog verhindert. Diese Reflexionen resümierend, solle der transitorische Aspekt in der Vorlage noch gestärkt werden.

Neben der Kontinuitätsfrage stellte sich auch die Frage nach dem Sinn des Vorschlags des hermeneutischen Annotierens, zumal das intuitive Annotieren die bisher bewährte Logik gefährde und auf Kosten der Maschinenlesbarkeit gehe. Die Trennung von beidem habe sich bislang als sinnvoll erwiesen, etwa wenn es um das Schärfen von Kategorien gehe. Hierzu wurde jedoch zu bedenken gegeben, dass diese skizzierte Opposition nicht unbedingt Bestand habe, da Vorannahmen immer schon einen wichtigen Bestandteil der Hermeneutik ausmachten. Bei der Dichotomie handele es sich daher eher um ein stetiges Kontinuum. Der Versuch

einer Unterscheidung führe folglich zu einem ‚Moving Target‘. Die Idee einer starken kategorialen Trennung sei eigentlich ein Artefakt – und ihrerseits historisch.

Eine Gretchenfrage sah die Diskussion darin, welche der im Laufe des Annotationsprozesses generierten Forschungsdaten behalten werden sollten. Der Autor des Beitrags ging hier sogar noch einen Schritt weiter, indem er in dieser Frage eine Überlebensfrage und Kernaufgabe der *Digital Humanities* sah: Lange Latenzzeiten seien nicht mehr zeitgemäß und alles, was Forschende leisteten, müsse als Prozess transparent gemacht werden. Die Forschungsdaten sollten viel schneller an die Öffentlichkeit gelangen und Nutzbarkeit müsse ermöglicht werden.

Zu weniger konkreten Ergebnissen führten Überlegungen bezüglich der Kenntnisse, die für die Benutzung von Plattformen vorausgesetzt werden sollen. So wurde der Wunsch nach einer möglichst niedrigschwelligen, ohne ‚DH-Kenntnisse‘ zu bedienenden Plattform kritisch gesehen. Eine Bedienbarkeit ohne jegliche Vorkenntnisse zur Textmodellierung sei weder realistisch noch wünschenswert. Auch die Frage nach Kenntnissen, die sinnvoll vorausgesetzt werden könnten, wurde zurückgewiesen mit dem Hinweis auf eine nicht bestimm- bare Menge von Vorkenntnissen, die zudem ständiger Veränderung unterliegen würden. Insgesamt sei jedoch ein sich stetig verbesserndes Niveau digitaler Kompetenzen zu beobachten, bei dem viele digitale Konzepte vorausgesetzt werden könnten.

Abschließend wurden Auswirkungen der im Beitrag und seiner Diskussion erzielten Erkenntnisse und Forderungen für das wissenschaftliche Feld besprochen. Hervorgehoben wurde hier die Integration in die Lehre, in der bei ausreichender Komplexität und Innovativität annotierte Texte als Leistungen zu akzeptieren seien. Ein mögliches Vorgehen könne darin bestehen, dass ein Anteil der Arbeit eine Methodenreflexion darstelle, in der grundlegende Konzepte der Annotation, ihre Reichweite, Fehlerhaftigkeit etc., thematisiert würden. Bisherige Erfahrungen zeigten, dass Studierende dieses Vorgehen sehr gern annehmen würden. Ein positiver Effekt, der auch traditioneller eingestellte KollegInnenen zu überzeugen vermöge, liege zudem darin, dass diese Arbeiten dazu führten, dass Texte genauer gelesen würden als in anderen Seminaren.

Der Beitrag von Evelyn Gius trägt den Titel *Digitale Hermeneutik. Computergestütztes Close Reading als traditionelles Forschungsparadigma?* und widmete sich somit dem Kerngeschäft der philologischen Analyse. Der Modus der Digitalität trage demnach nicht nur dazu bei, ‚dunkle Flecken‘ der Texterschließung zu erhellen, er führe auch zur Offenlegung des Analyseprozesses, der sehr viel stärker ins Bewusstsein rücke und in dem Präzision und kollaborative Verfahren einen hohen Stellenwert einnehmen würden.

Vor diesem Hintergrund ging es in der Diskussion um das Verständnis von Hermeneutik, unter der mehr zu fassen sei als das Verstehen beim digitalen Annotieren. Dazu gehörten etwa die systematische Einordnung in ein System mit weiteren Texten, die Neu-Perspektivierung und Verfremdungsprozesse. Es stelle sich daher die Frage, inwiefern die hermeneutische Begriffsvielfalt für eine

digitale Hermeneutik reduziert werden müsse. Hierzu erläuterte die Autorin, dass der im Beitrag entworfene Zugang keine neue Hermeneutik anstrebe, sondern lediglich dazu dienen solle, Textanalyseprozesse genauer zu betrachten. Es gehe daher nicht um eine neue Systematik, sondern um die Möglichkeit zu einer kontrastiven Betrachtung. Dabei wollten sich die *Digital Humanities* keinesfalls als eine Substituierung verstanden wissen, hätten in einigen Bereichen vielmehr Angebote vorzuweisen.

Ein weiterer Kommentar zur Terminologie betraf die Unschärfe der Begriffe *Close Reading* und *Hermeneutik*. Auf beide werde zurückgegriffen, um einen Blick auf nicht-digitale Verfahren zu werfen, was dazu führe, die dazwischenliegende Kluft zu unterschätzen, da jeweils nicht eindeutig zu bestimmen sei, was darunter gefasst werde. Anknüpfend an eine in den vorausgehenden Diskussionen bereits kommunizierte Kritik am Begriff des *Close Reading* konkretisierte Gius ihre Lesart als ein stellenbezogenes Lesen und Kommunizieren und fügte an, dass sie nicht auf das Festhalten dieser Begriffe bestehe. Eine präzisere Terminologie stehe derzeit jedoch noch aus.

In Bezug auf die Hermeneutik schloss sich die Frage an, inwiefern ästhetische Erfahrungen, die hier traditionell stark einfließen, im Digitalen im Grunde gelöscht würden und eine ‚Anästhetisierung des Gegenstands‘ bewirkten. Die Autorin sah in digitalen Verfahren nicht unbedingt eine solche Anästhetisierung, sondern vielmehr eine Verlagerung des Ästhetischen auf andere Ebenen. In der Diskussion wurde zudem angemerkt, dass es – auch in anderen Disziplinen – seit jeher Spannungen zwischen ästhetischen und formalisierenden Verfahren gebe, und exemplarisch auf einen Aufsatz von Christoph März mit dem Titel *Metrik, eine Wissenschaft zwischen Zählen und Schwärmen?* (1999) verwiesen. Auch sei die Formalisierbarkeit ästhetischer Erfahrungen insofern vorstellbar, als beispielsweise Metaphern oder Metonymien annotierbar und somit in der Lage seien, ästhetische Erfahrungen anzuzeigen oder einzuleiten.

Neben den hermeneutisch-terminologischen Aspekten widmete sich die Diskussion ausführlich der Korpusbildung. Hier bestehe ein Problem des Zirkulären, da manuell aufbereitete Korpora bereits eines gewissen Vorverständnisses bedürften, dieses jedoch Einfluss nehme auf darauf basierende Ergebnisse. Das Problem bestehe zwar auch in anderen Disziplinen, linguistische Korpora könnten es jedoch bereits durch Quantität nivellieren. Entsprechende Ansätze sieht Gius auch für die Literaturwissenschaft, beispielsweise, wenn aus einer möglichst großen Textmenge eines für die Analyse relevanten Zeitraums automatisiert relevante Texte herausgesucht würden, anhand derer entsprechende Operationalisierungen vorgenommen werden könnten. Dass es immer auch ein Zirkel bleibe, der qualitative und quantitative Verfahren miteinander vereine, bedeute auch, dass Analyse ein Ausgangspunkt für Heuristik sein könne – was aber nicht problematisch sein müsse. Verkürzt und ein wenig vorsichtiger formulieren ließe sich dieser Sachverhalt mit einer Gleichsetzung von Analyse- und Modellierungsfragen.

Im Anschluss an die Frage zum Methodenmix wurde die im Beitrag vorgenommene scharfe Trennung von digitalen und nicht-digitalen Verfahren kritisch

hinterfragt. Eine solche Unterscheidung entspreche nicht der Realität, da man *Digitale Geisteswissenschaften* prinzipiell auch analog betreiben könne. Folglich sei eine Unterscheidung in formalisierende und nicht-formalisierende Ansätze passender. Zunächst wurde hierzu angemerkt, dass auch Unsystematisches annotiert und formalisiert werden könne. Unter strategischen Gesichtspunkten müsse man sich die Frage nach der Verortung der *Digital Humanities* stellen. Man stehe sehr stark in etablierten Traditionen. Annotationsprojekte stellten dabei ein sehr gutes Gesprächsangebot dar, indem sie einen Ansatz böten, der eine Brückenfunktion einnehmen könne, insbesondere wenn sie ihre Aus-handlungsprozesse sichtbar hinterlegten. Als zentraler Bestandteil rückte hier der Interpretationsbegriff in den Fokus. Während die Prozessreflexion und die Kategorialreflexion des Beitrags gelobt wurden, machte man darauf aufmerksam, dass eine Präzisierung des Interpretationsbegriffes dazu führen könne, die literaturwissenschaftliche Provenienz besser einzubeziehen. Es müsse klar werden, was Interpretieren tue, wenn sie interpretieren. Die extreme Vielfalt und Heterogenität von Interpretationen müsse hierzu kategorial erfasst und mit den jeweiligen Zielen assoziiert werden. Eine Schwierigkeit bei dieser wissenschaftlichen Selbstreflexion bestehe jedoch in einer Leerstelle des Experiments, nämlich der, die das Vorgehen beschreibe, das angewandt wird, wenn keine Tools zur Bearbeitung derselben Fragestellung eingesetzt würden.

Abschließend wurde in diesem Zusammenhang diskutiert, inwiefern man sich in Fragen der disziplinären Selbstbestimmung und Kartographierung von anderen Fächern inspirieren lassen könne, in denen die Verankerung von Tools bereits weiter fortgeschritten sei. In den Sozialwissenschaften, beispielsweise, sei man in dem Anlegen von Interpretationsvorstufen viel weiter, weshalb sie sich als Impulsgeber für Prozessreflexionen eignen könnten. Einschränkend sei hier jedoch anzufügen, dass sich derartige Ansätze womöglich für das Identifizieren vergleichbarer Prozesse eigneten, für deren Reflexion jedoch untauglich seien, da in unterschiedlichen Disziplinen andere Prämissen vorzufinden seien.

Mit dem Beitrag von Rüdiger Nutt-Kofoth schloss sich der Kreis der Sektion, indem er genuin editionsphilologische Probleme tangierte. Er unternahm eine differenzierte und kritische Abwägung digitaler und analoger Editionen mittels systematischer Vergleiche inhaltlicher und funktionaler Kategorien. Aus den Beobachtungen ergab sich hier eine gewisse Relativierung des ‚Neuen‘, was die im Digitalen behauptete Qualitätsinnovation betrifft. Die Tendenz zum Aufbau von digitalen Wissensverbänden mit archivalischen Zügen habe zudem zur Folge, dass die Betitelung als *historisch-kritische Ausgabe* und überhaupt eine differenzierende Typologie im Digitalen nur noch selten auftrete – die *Faust*-Edition bilde hier eine Ausnahme.

Das Programm der digitalen historischen Edition reflektierend, unternahm die Diskussion den Versuch, ihre Grenzen auszuloten. Dabei stellte sich die Frage, ob es notwendig sei, die Volltexte aller Quellen beizufügen – wohlwissend, dass es sich hierbei um einen substanziellen Teil der abendländischen Literatur handeln könne. Bei der Alternative, der Ergänzung des Angebots über externe Verlinkungen, bestehe das Problem eines erhöhten Wartungsaufwandes. Während

sich an dieser Zustandsbeschreibung der erhöhte Bedarf nach Standardisierungen zeige, müsse der *Digitalen Edition* im gegenwärtigen ‚Inkunabelzeitalter‘ womöglich eine gewisse Experimentierfreude zugestanden werden, um das skizzierte Spannungsfeld auszubalancieren.

Ähnliches gilt für die Bewertung des Phänomens ‚Medienwandel‘. Anknüpfend an die Überlegungen im Rahmen des Beitrags von Braun, Viehhauser und Zirker, wurde es unter revolutionären und evolutionären Gesichtspunkten diskutiert. Bei der ‚Revolution‘ handle es sich um einen Umbruch von Ordnungen, der auf den Transfer der Edition nicht zuzutreffen scheine. Ohne sich an dem Begriffspaar völlig abarbeiten zu wollen, wurde auf den von Elena Pierazzo geprägten Begriff der *Radical Evolution* verwiesen, der möglicherweise zutreffend sein könne. Allerdings müssten für eine Bewertung eine belastbare Menge von Durchschnittsverwendungen von Editionen betrachtet werden. Bislang bestimmten einzelne größere Editionsprojekte die Diskussion, es gebe jedoch gar nicht so viele neugermanistische *Digitale Editionen*, die jetzt schon wissenschaftsgeschichtlich beschrieben werden könnten. Für das Identifizieren von Ausreißern und damit das Eintreten eines Bruchs sei zunächst eine solche repräsentative Menge erforderlich.

Anschließend wurde darauf hingewiesen, dass das Festhalten am Konzept der Edition auch in ihrem medialen Transfer ein wesentliches Merkmal bleiben müsse. Der Begriff des ‚Archivs‘, der den Medienwandel der *Digitalen Edition* begleite, verschleierte beispielsweise das identitätsstiftende Objekt der Edition, indem er sie auf etwas abstrakt Virtuelles beschränke. Eine wichtige Rolle komme hierbei dem Interface zu, das auf funktionaler Ebene seine Konzeptualität transportiere. Es wäre wünschenswert, so die Rückmeldung, wenn dieser Aspekt im Beitrag noch gestärkt werden könnte. Das Überführen der etablierten editorischen Kriterien in das digitale Paradigma biete großes Potenzial.

Einen qualitativen Unterschied im Vergleich zur traditionellen Printedition sah der Autor des Beitrags auch in den Möglichkeiten der Multimedialität. Mit dem Verweis auf deren bereits in Printeditionen erfolgte Einbindung – etwa durch die Beigabe von CD-ROMs – wurde der Status dieses Arguments hinterfragt. Die anschließende Diskussion machte deutlich, dass im Digitalen keine reine Medienkombinatorik gemeint sei, sondern ein Großzusammenhang hergestellt werde, der einfachere Zugänge für Medienwechsel biete und ‚offene Grenzen‘ fördere. So könnten beispielsweise Werke von AutorInnen, die in mehreren unterschiedlichen Medien gewirkt haben, miteinander verknüpft werden. In der Transmedialität bestehe ein wesentliches Spezifikum der *Digitalen Edition*. Auch hier zeige der Großzusammenhang, dass quantitative Argumente in qualitative umschlagen könnten, sobald sie durch entsprechende Konzepte in diesen eingebettet werden.

Vor dem Hintergrund einer derartigen Ausweitung des Angebots der *Digitalen Edition* widmete sich die Diskussion der Perspektive der RezipientInnen und den Konsequenzen, die dieses Forschungsdesign für deren Lektüerverhalten hat. Obwohl hier sicherlich erhebliche Veränderungen aufträten, die auch das Verhältnis zwischen NutzerInnen und HerausgeberInnen betreffen, könne hier von einem zunehmend wachsenden Medienvorwissen gerade jüngerer RezipientInnenen ausgegangen werden. Darüber hinaus werde nicht genug über die Oberflächen

gesprochen, die Nutzenden den Zugriff auf die Edition erlauben. Ihre Gestaltung und intuitive Bedienbarkeit würden noch zu häufig nicht rechtzeitig in laufende Editionsprozesse eingebunden, was im Nachhinein zu einem erhöhten Aufwand führe. Insbesondere aufgrund der Komplexität der zur Verfügung gestellten Daten sei es auch Aufgabe der Oberfläche, mögliche Wege durch das Textuniversum aufzuzeigen und verschiedene Zugänge anzubieten. Es handle sich bei einer Oberfläche um ein ‚Entscheidungszugangsphänomen‘. Gleichzeitig liege die Neuartigkeit der *Digitalen Edition* jedoch auch in der Möglichkeit, die hinter den Editionen liegenden Daten als solche zu betrachten und zu entkontextualisieren – mit der Perspektive, sie auch wieder zurückspielen zu können. Bei *Digitalen Editionen* handle es sich folglich um *Editions as Interfaces*, deren Ausstattung mit einer API eine weitere *Datafication* und somit unterschiedliche Zugänge für heterogene Nutzergruppen erlaube. Es seien die sich hier eröffnenden Anschlusspotenziale, die Spezifika der *Digitalen Edition* ausmachen. Hier fehle es gegenwärtig noch an Beispielen aus realisierten Editionen, anhand derer die Brücke zwischen Edition und Anschlussoperation sichtbar gemacht werden könnte. Die Verlässlichkeit der Edition und die in ihr erbrachte textkritische Leistung müsse weiterhin als bewährter Standard erhalten bleiben und transparent gemacht werden. Folglich müssten editorische Grenzen gezogen und gleichzeitig Wege gefunden werden, um individuelle Anschlussoperationen zu ermöglichen.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.



Sektion III. Textanalyse

Einführung

Jan Christoph Meister

Spätestens seit dem Strukturalismus gilt die Befassung mit formalen Eigenschaften literarischer Texte nicht nur für die literaturwissenschaftliche Befassung mit lyrischen Texten als ein methodisches *sine qua non*. Auch in Hinblick auf Texte der dramatischen und epischen Gattung ruht eine wissenschaftliche, rational argumentierenden Auseinandersetzung nach dem gegenwärtigen Selbstverständnis moderner Literaturwissenschaften, wo sich diese als Textwissenschaften definieren, mittlerweile notwendig auf einer methodisch ausgewiesenen Textanalyse auf.

Unter den Begriff „Textanalyse“ fallen dabei allerdings zweierlei Dinge – die Textanalyse als Textsorte, und die Textanalyse als geregelttes Verfahren. Der Fokus der gegenwärtigen Sektion liegt auf dem Verfahrensaspekt. Was aber ist ein textanalytisches Verfahren, und warum ist eine Verständigung über das Konzept im Rahmen einer digitalen Literaturwissenschaft angezeigt? Dazu seien einleitend drei Thesen formuliert:

1. Was unter den Begriff „Analyse“ – und also auch: Textanalyse – fällt, lässt sich sowohl unter dem Gesichtspunkt der argumentativen & strategischen Funktion wie unter dem der methodischen Organisation betrachten. Ich möchte hier die These wagen, dass wir im wissenschaftlichen Diskurs der digitalen Literaturwissenschaften gegenwärtig dazu tendieren, umstandslos auf die Organisation und die Verfahrenslogik textanalytischer Verfahren im Sinne angewandter Methoden zu fokussieren. Das ist im Sinne der Optimierung und Professionalisierung der digitalen Literaturwissenschaften zwar gut und wichtig – aber wir sollten diese Verfahren durchaus auch unter

J. C. Meister (✉)
Universität Hamburg, Hamburg, Deutschland
E-Mail: jan-c-meister@uni-hamburg.de

dem Gesichtspunkt ihres rhetorischen und institutionellen Profils innerhalb des methodischen Gesamtspektrums der Textwissenschaften befragen. Auch und gerade formale Gegenstandsanalysen operieren immer vor dem Hintergrund einer Gegenstandstheorie, zu der sie sich konzeptionell affirmativ oder kritisch positionieren; sie fungieren zudem oftmals als das zentrale interdisziplinäre Interface, über das nicht allein Konzepte und Verfahren ausgetauscht werden, sondern auch die unterschiedlichen disziplinären Praktiken der sozialen Organisation von Forschung untereinander vermittelt werden können.

2. Eine der spannendsten Fragen im Zusammenhang der literaturwissenschaftlichen Methodenreflexion ist die Möglichkeit zur Unterscheidung zwischen explorativen und hypothesengeleiteten, experimentellen Textanalysen. Indem die digitale Literaturwissenschaft sich dieser Frage widmet, forciert sie nun allerdings durchaus keinen ‚Paradigmenwechsel‘, im Gegenteil: sie schreibt eine weitaus ältere Tradition fort. So lesen wir etwa in Jochen Vogts weit verbreiteter *Einladung zur Literaturwissenschaft* aus dem Jahr 1999 zum Thema *Hermeneutik*:

Sie wird einerseits, den hermeneutischen Regeln von Schleiermacher bis Habermas folgend, den Sachgehalt ihrer Texte im Kommentar erschließen. Sie muss aber andererseits immer die prinzipielle Vieldeutigkeit literarischer Texte bedenken und aus ihren jeweiligen Kommentaren nur behutsame Deutungen ableiten. Jede einzelne muss sich dabei um Plausibilität und Nachvollziehbarkeit bemühen, d. h. möglichst viele Aspekte eines gegebenen Textes integrieren und dabei anderen, konkurrierenden Deutungen standhalten können. Vor allem aber wird diese literarische Hermeneutik darauf achten müssen, durch welche Kunstmittel ein Text Bedeutung, besser: Bedeutungen, Sinn-Potentiale erzeugt.¹

Was Vogt hier mit Blick auf die hermeneutische Texterschließung hervorhebt, ist die erforderliche Sensibilität für die textuellen, semantisch produktiven „Kunstmittel“. Einige dieser ‚Kunstmittel‘ lassen sich heute mittels digitaler Verfahren durchaus im Zuge eines textanalytischen *preprocessing* fassen. Nur: viele dieser digitalen Verfahren greifen dabei auf Modellierungen von Textphänomenen zurück, die aus der Linguistik, der kognitivistisch orientierten Informatik oder der mathematisch-statistischen Stochastik stammen. An die Stelle der expliziten, inhaltsbezogenen Deutungshypothese traditioneller Textauslegungen tritt so im Zuge der digitalen Operationalisierung literaturwissenschaftlicher Konzepte eine implizite Modellhypothese. Diese Modell- und Modellierungshypothesen kritisch zu reflektieren und sich darüber Klarheit zu verschaffen, inwiefern sie möglicherweise das jeweils thematische Textphänomen und ‚Kunstmittel‘ konzeptionell verfremden, ist für die digitalen Literaturwissenschaften von zentraler Bedeutung.

¹Jochen Vogt: *Einladung zur Literaturwissenschaft*. Paderborn (Fink) 2008. Hier zitiert nach der 7., erweiterten und aktualisierten Auflage 2016, S. 69.

3. Das Konzept ‚Analyse‘ ist per se nicht selbstexplikativ. Im einleitenden Abschnitt des Eintrags der *Stanford Encyclopedia of Philosophy* zu dem Lemma lesen wir dazu:

If asked what ‘analysis’ means, most people today immediately think of breaking something down into its components; and this is how analysis tends to be officially characterized. (...) This conception may be called the *decompositional* conception of analysis (...) But it is not the only conception, and indeed is arguably neither the dominant conception in the pre-modern period nor the conception that is characteristic of at least one major strand in ‘analytic’ philosophy. In ancient Greek thought, ‘analysis’ referred primarily to the process of working back to first principles by means of which something could then be demonstrated. This conception may be called the *regressive* conception of analysis (...). In the work of Frege and Russell, on the other hand, before the process of decomposition could take place, the statements to be analyzed had first to be translated into their ‘correct’ logical form (...). This suggests that analysis also involves a *transformative* or *interpretive* dimension.²

Bei der Lektüre der Beiträge in der Sektion „Textanalyse“ bleibt diese wichtige Unterscheidung im Hinterkopf zu behalten: wir sollten die im einzelnen vorgestellten Verfahren also darauf befragen, ob sie auf eine Textanalyse im Sinne der (wohlgemerkt logisch, nicht generativ oder genealogisch gemeinten) *dekompositionellen*, der *regressiven* oder der *transformativen* bzw. *interpretativen* Analysekonzeption hinauslaufen. Insbesondere der Hinweis auf den transformationellen, interpretativen Effekt der logischen Formalisierung, die der analytischen Dekomposition vorausgeht, scheint mir dabei einer zu sein, der für eine reflektierte Praxis wie für die weitere Methodenentwicklung im Feld der digitalen Textanalyse wichtig ist.

Damit zu den Beiträgen in dieser Sektion. In den nachfolgend präsentierten Überlegungen zu einer digitalen Textanalyse begegnen uns vier Konstellationen von Gegenstand und Methode:

- erstens die Anwendung von an Korpora entwickelten bzw. trainierten digitalen textanalytischen Verfahren auf Einzeltexte, deren Profilierung oder Strukturierung analysiert werden soll (Berenike Hermann: *Operationalisierung der Metapher zur quantifizierenden Untersuchung deutschsprachiger Texte im Übergang vom Realismus zur Moderne*);
- zweitens die korpusanalytische Textanalyse, die summarische oder differentielle Befunde zur untersuchten Textgesamtheit erarbeitet und dabei u. a. auch kanonisiertes ‚Expertenwissen‘ auf den Prüfstand stellt (Marcus Willand & Nils Reiter: *Quantitative Analyse zur thematischen*

²Beaney, Michael, „Analysis“, The Stanford Encyclopedia of Philosophy (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2018/entries/analysis/>>.

Figurenkonzeption im Drama; Thomas Weitin: Distanzmaße und Netzwerkanalysen bei einem mittelgroßen literarischen Korpus);

- drittens die Erprobung und Diskussion neuer quantitativer und statistischer Verfahren, die als einzelne (Christof Schöch: *Quantitative Semantik: word2vec für literaturwissenschaftliche Fragestellungen*) oder in Kombination (Peer Trilcke: *Small Worlds, Change Rates und die Netzwerkanalyse dramatischer Texte*) als Kandidaten für das digitale literaturwissenschaftliche Methodenrepertoire infrage kommen;
- viertens die Selektion und partielle Abbildung dieses Repertoires, wie sie in der Konzeptionierungsphase z. B. eines DH-Projektes zur historischen Korpusanalyse erforderlich ist (Maciej Maryl: *Operationalising the Change: Theoretical Foundations for the Quantitative History of Polish Literature 1989–2000*).

Auffällig ist bei dieser Gesamtschau, dass das Thema „digitale Textanalyse“ zwar nicht ausschließlich, aber doch überwiegend korpusbezogen verhandelt wird. Das hat einerseits plausible methodologische und technische Gründe, z. B. dort, wo statistische Verfahren involviert sind. Diese „korpuszentrierte“ Orientierung unseres Nachdenkens über die digitale Textanalyse wirft aber zugleich auch eine methodologische Grundsatzfrage auf, die nicht zuletzt strategische Konsequenzen für die Positionierung der digitalen Literaturwissenschaften vis-à-vis den traditionellen Philologien haben könnte. Zugespißt und bewusst polemisch gefragt lautet ein möglicher Einwand: Haben wir uns in der textanalytischen Praxis der digitalen Literaturwissenschaften faktisch bereits von der Kategorie des Einzeltextes als überhaupt legitimem Gegenstand, an dem man etwas lernen kann, verabschiedet? Können in der digitalen Literaturwissenschaft die in der traditionellen Philologien im Zentrum der Aufmerksamkeit stehenden Einzeltexte von uns aus methodischer Notwendigkeit nur noch differentiell bzw. relational analysiert werden?

Wenn dem tatsächlich so sein sollte, dann hieße dies, dass der Einzeltext für die Weiterentwicklung digitaler Methoden der Textanalyse eine mittlerweile wegzuführende epistemische Leiter geworden ist und insofern einer Gegenstandsklasse angehört, über die wir nicht mehr reden können, sondern schweigen müssen. Im Nachdenken und Reden über das Thema „digitale Textanalyse“ ist nicht zuletzt zu dieser Frage eine klare, explizite Positionierung anzustreben – denn sie ist eine, die viele Vertreter der traditionellen berechtigter Weise an die digitale Literaturwissenschaft richten.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





What are they Talking About? A Systematic Exploration of Theme Identification Methods for Character Speech in Dramatic Texts

Nils Reiter und Marcus Willand

1 Introduction

This article investigates theme detection in character speech of dramatic texts. It does not aim to contribute to drama(-historical) research, but introduces two methodological innovations to computational literary studies: We propose a procedure for the systematic comparison of text-analytic methods even if a gold standard does not exist or is difficult to create. We employ this procedure for the comparison of several implemented systems that assign some notion of ‘theme’ to a text segment. Henceforth, we use the term ‘theme’ to describe the content of character speech and the term ‘topic’ exclusively for the outcome of topic modeling. ‘Theme’ represents the goal, and ‘topic’ the outcome of one specific method (that may or may not achieve this goal).

The lack of reference corpora with annotated categories that are of relevance for literary studies is a methodological bottleneck for the development of text processing tools for literary texts. Existing and publicly available corpora allow researchers to test and/or train their tools, and to assess their quality. The existence of such corpora in other domains and with non-literary categories has been the driving force behind the progress made in computational linguistics over the past years. Creating reference corpora for literary phenomena, however, is a

N. Reiter (✉)
Universität zu Köln, Köln, Deutschland
E-Mail: nils.reiter@uni-koeln.de

M. Willand
Heidelberg, Deutschland
E-Mail: marcus.willand@gs.uni-heidelberg.de

challenging task, for multiple reasons.¹ One reason is the nature of literary texts, as they are often deliberately ambiguous and open for different interpretations. In addition, literary categories are hard to define exactly,² which is often only uncovered during annotation.³ In this article, we would like to offer a different perspective on the issue: Knowledge that has been established in literary studies might not be defined as exact as in natural science, but with the right procedure it can still be employed for method development.

The procedure we propose can be outlined as follows: We first formulate a set of postulates about dramatic texts. These postulates combine a structural element of the dramatic text with a non-interpretive hypothesis about the thematic content of this text element. For instance, we postulate that continuous copresence of characters leads to similar thematic content of their dramatic speech. The procedure does not require these postulates to be true for every drama. Though, we *do* maintain that they are plausible assumptions for a large number of dramatic texts and that results obtained by tuning or training on these postulates would be suitable for poetological or drama historical statements.

Dramatic texts are well suited for this analysis, because they are strongly structured by nature (compared to, e.g., prose texts) and this structure is encodable in a machine-readable fashion. The speech of characters is unfiltered by a narrator, the entire text is segmented into acts and scenes. Unsurprisingly, the dramatic structure has always played a role in the literary analysis of dramas. Many theoretical predictions are rooted in the drama structure that makes both the formulation of general postulates and their operationalization straightforward.

Most work in the past has been either analyzing the dramatic structure or the content of dramatic language. For social network analysis, one typically extracts co-presence information (which characters share the stage how often?) from the texts.⁴ On top of the relational information encoded in networks, Algee-Hewitt

¹Marcus Willand, "Hermeneutische Interpretation und digitale Analyse. Eine Verhältnisbestimmung," in: Luisa Banki (ed.), *Lektüren. Positionen Zeitgenössischer Philologie*, Trier 2017, 77–100.

²Cf. Nils Reiter/Evelyn Gius/Jannik Strötgen et al., "A Shared Task for a Shared Goal – Systematic Annotation of Literary Texts," in: *Digital Humanities 2017: Conference Abstracts*, Montreal 2017, <https://dh2017.adho.org/abstracts/192/192.pdf> (retrieved on June 10, 2018).

³Thomas Bögel/Michael Gertz/Evelyn Gius et al., "Collaborative Text Annotation Meets Machine Learning: heureCLÉA, a Digital Heuristic of Narrative," in: *DHCommons 1* (July 2015), <https://doi.org/10.5281/zenodo.3240591>.

; Bernhard Fisseni/Aadil Kurji/Benedikt Löwe, "Annotating with Propp's Morphology of the Folktale: Reproducibility and Trainability," in: *Literary and Linguistic Computing* 29/4 (2014), 488–510, [10.1093/lc/fqu050](https://doi.org/10.1093/lc/fqu050) (both retrieved on June 10, 2018).

⁴Franco Moretti, "Network Theory, Plot Analysis," in: *Pamphlets of the Stanford Literary Lab* 2 (2011), <https://litlab.stanford.edu/LiteraryLabPamphlet2.pdf>; Peer Trilcke/Frank Fischer/Dario Kampkaspar, "Digital Network Analysis of Dramatic Texts," in: *DH2015 Conference Abstracts*, Sydney 2015, <https://dlina.github.io/presentations/2015-sydney/sydney.html#/> (both retrieved on June 10, 2018).

uses betweenness centrality to make interesting observations.⁵ Genre analysis, as for instance done by Schöch,⁶ on the other hand, ignores most of the structure and just takes the text as a whole: All characters' voices are merged into one.

There are only a few works that actually combine structure and content information: Nalisnick/Baird have employed sentiment analysis on the character speech to determine the turning point in the relation between characters (e.g., between Hamlet and his mother).⁷ Bullard/Ovesdotter Alm describe corpus-linguistic investigations into two Scandinavian authors and determine how certain character classes are represented linguistically.⁸ Karsdorp et al. focus on one specific character relation and describe a system to determine the most probable romantic partner for characters.⁹

This paper is structured in the following way: In Sect. 2, we discuss conceptually how we are going to compare systems on a task for which a gold standard does not exist. Section 3 introduces the theme detection systems. In Sect. 4 we discuss four postulates about dramatic texts, against which we compare the systems. Section 5 outlines the experimental setup and introduces formal and notational basics. In Sect. 6, each postulate is covered by one subsection, discussing its operationalization and experimental results. Section 7 concludes.

2 Comparing Systems without a Gold Standard

In order to compare the performance of different text analysis systems, one would typically presuppose the existence of a gold standard, against which the systems can be tested. While having a gold standard would be ideal, creating one is difficult in this case for two reasons that are related with the target phenomenon 'theme'.

⁵Mark Algee-Hewitt, "Distributed Character: Quantitative Models of the English Stage, 1500–1920," in: *Book of abstracts* (2017), 119–121, <https://dh2017.adho.org/abstracts/DH2017-abstracts.pdf> (retrieved on June 10, 2018).

⁶Christof Schöch, "Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama," in: *Digital Humanities Quarterly* 2 (November 2017), 10.5281/zenodo.166356.

⁷Eric T.Nalisnick/Henry S. Baird, "Character-to-Character Sentiment Analysis in Shakespeare's Plays," in: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Vol. 2: *Short Papers*, Sofia 2013, 479–483, <http://www.aclweb.org/anthology/P13-2085> (retrieved on June 10, 2018).

⁸Joseph Bullard/Cecilia Ovesdotter Alm, "Computational Analysis to Explore Authors' Depiction of Characters," in: *Proceedings of the 3rd Workshop on Computational Linguistics for Literature* (2014), 11–16, <http://www.aclweb.org/anthology/W14-0902> (retrieved on June 10, 2018).

⁹Folger Karsdorp/Mike Kestemont/Christof Schöch et al., "The Love Equation: Computational Modeling of Romantic Relationships in French Classical Drama," in: Mark A. Finlayson/Antonio Lieto/Ben Miller (eds.), *6th Workshop on Computational Models of Narrative* 45 (2015), 98–107, 10.4230/OASIS.CMN.2015.98.

Defining a list of themes as annotation categories would in principle allow us to annotate text segments (e.g., character utterances) with such categories. Cast as a supervised classification problem, it would be straightforward to train a classifier that detects and distinguishes these themes. During the application of this classifier, however, we would only be able to detect the topics as they are defined in the gold standard. We are then not able to discover new themes in the corpus, in addition to the usual dependency on the training data. The requirement of seeing negative data during training (i.e., texts that do *not* refer to a given theme) makes even binary classification attempts (training one classifier for one theme) difficult, as the *absence* of a theme may be even harder to annotate than its presence. All this is also related to another challenge: Themes in literary texts are to some extent interpretative, in particular if the theme is deliberately hidden or its detection depends on special historical knowledge. Except for very broad and general themes, it is therefore unlikely that any kind of gold standard will be created in the near future.

The core idea in this procedure is to formulate a number of postulates that are derived from existing scholarly knowledge about dramatic texts. These postulates are assumed to be true *in general* without having to be true for every single text. Such postulates do need to be operationalized at some point, and structural properties of both the systems in comparison and the postulates make some more suited than others: Postulates that state relations between one variable and another can be operationalized as correlations in a straightforward manner, for instance. Once this is done, the different theme detection systems can be applied onto a corpus and their output be compared with the postulates. However, the interpretation of such results cannot be expected to be trivial, as we are comparing multiple systems with multiple postulates.

Conceptually, this is similar to two existing concepts from *natural language processing (NLP)*: In the paradigm of distant supervision, an un-annotated corpus is annotated automatically, based on structured data that is in some way linked to the corpus. In Husby/Barbosa, training data is created by linking blog articles to *Wikipedia* domains via freebase entities,¹⁰ which are in turn used as topic labels for the blog articles. After the data has been established, standard supervised classification algorithms can be applied. The second concept is extrinsic evaluation. NLP tools are often used in pipelines: The output of a part of speech tagger is fed as input to a dependency parser (the downstream task). If there was no gold standard for part of speech tags, one could evaluate a part of speech tagger by comparing the performance of a dependency parser with and without the input of the part of speech tagger (assuming we have an evaluation method for the dependency parser).

If formalized in an appropriate way, existing literary knowledge can actually be re-used as a '*distant gold standard*'. Systems could be optimized to reproduce decisions and distinctions made by literary scholars.

¹⁰Stephanie Husby/Denilson Barbosa, "Topic Classification of Blog Posts Using Distant Supervision," in: *Proceedings of the Workshop on Semantic Analysis in Social Media 2012*, 28–36, <http://www.aclweb.org/anthology/W12-0604> (retrieved on June 10, 2018).

Table 1 Overview of the theme detection systems. A detailed description of the columns can be found in the text

Method	Dimensions	Word classes	Average size	Total size
Custom dictionary	5	content	88.8 (± 23.4)	444
Enriched custom dictionary	5	content	178.6 (± 32.6)	893
<i>GermaNet</i> dictionary	54	content	2,644.7 ($\pm 3,837.9$)	142,815
TM by drama (k = 20)	20	all	22,798.16 ($\pm 10,389.32$)	159
TM by character (k = 20)	20	all	819 ($\pm 1,650.5$)	4,426
TM by drama (k = 100)	100	all	22,798.16 ($\pm 10,389.32$)	159
TM by character (k = 100)	100	all	819 ($\pm 1,650.5$)	4,426

3 Theme Detection Systems

We will be using two kinds of systems in this article: Dictionary-based systems use a dictionary to detect words in text that belong to certain keywords in this dictionary. Dictionary-based systems require the existence or creation of a dictionary, which pre-determines the themes one is able to find in texts. While defined themes might still appear in unexpected places, it is impossible to detect completely new themes. The findings, however, can be interpreted in a straightforward way, as resulting numbers represent the (relative) amount of words belonging to a certain keyword in a text segment.

Latent Dirichlet Allocation (LDA, referred to here as „topic modeling“, TM) is an unsupervised statistical method that aims at uncovering textual patterns by inspecting co-occurrence of words, typically in large corpora. Using topic models, one can discover new and unexpected themes in a corpus. However, because documents in topic models are represented as distributions over distributions over words (see below), they are difficult to interpret. As other researchers have discovered,¹¹ many of the resulting topics do not represent themes at all, but formal or structural properties of the documents. In addition, one has to specify the number of topics in advance.

To give an overview, Table 1 shows the systems and system variants in comparison. *Word classes* shows the set of parts of speech that have been taken into account for the analysis. The dictionary-based systems only work with content words (nouns, verbs, adjectives) and ignore function words. *Avg. size* shows the average size (and standard deviation) for keywords (dictionary methods)

¹¹E.g., Benjamin Schmidt, “Words Alone: Dismantling Topic Models in the Humanities,” in: *Journal of Digital Humanities* 2/1 (2012), <http://journalofdigitalhumanities.org/2-1/words-alone-by-benjamin-m-schmidt/> (retrieved on June 10, 2018).

or training documents (topic modeling). *Total size* gives the total number of entries (in the dictionaries) or number of training documents (for topic modeling).

3.1 Dictionary-based Approaches

A dictionary in this context is defined as a set of keywords with associated entries. The entries are semantically related to the keyword, but not necessarily synonymous. The keyword “theater”, for instance, would contain entries like “stage”, “actress”, “playing” etc. The dictionaries cover different word classes, historic, stylistic and spelling variations. They contain all entries as lemmas.

We will compare three different dictionary-based systems. The major difference between the systems lays in the nature and source of the entries. All systems assign a score for each keyword in the dictionary, based on the number of words that belong to this keyword. This score is normalized for the segment length, and, if word set sizes are differing largely, by dictionary size.

Custom Dictionary In the first variant of this method, we use a dictionary that has been collected by a domain expert (one of the authors of this paper) for the specific purpose of plays. As described in Willand/Reiter,¹² we manually created a dictionary for each of the five keywords that we consider exceptionally relevant for the analysis of dramatic texts from our research period (1730 to 1930): “family”, “war”, “love”, “reason” and “religion”. Each of the sets contains 60 to 115 words, which sum up to 444 entries in total. Time-span related meaning variation has been taken into account. The creation of the sets was done in an iterative manner: Starting from the word “love” we looked for words in its immediate context with the character speech that were then checked for their contexts.¹³

An illustrative example (taken from *ibid.*) on the use of these dictionaries is shown in Fig. 1. Each corner in the ‘spider web’ represents an entry in the dictionary field and each node represents the extent to which a character uses words that are grouped under this keyword. As Fig. 1 shows, a custom dictionary is capable of giving information about the thematic distribution of character speech in the dimensions defined in the dictionary. Apparently, the lovers mostly talk about love and the parents mostly talk about family (with respect to the five thematic fields). The crucial point of this method is that one has to have an idea of relevant themes before doing any analysis. This allows domain experts to incorporate their previous knowledge but has the downside that it limits what we can find to what we have defined before.

¹²Marcus Willand/Nils Reiter, “Geschlecht und Gattung. Digitale Analysen von Kleists *Familie Schrockenstein*,” in: *Kleist-Jahrbuch* (2017), 177–195, p. 184.

¹³The full dictionary is available on *GitHub*: <https://github.com/quadrana/metadata/tree/b5ab13/fields> (retrieved on June 10, 2018).

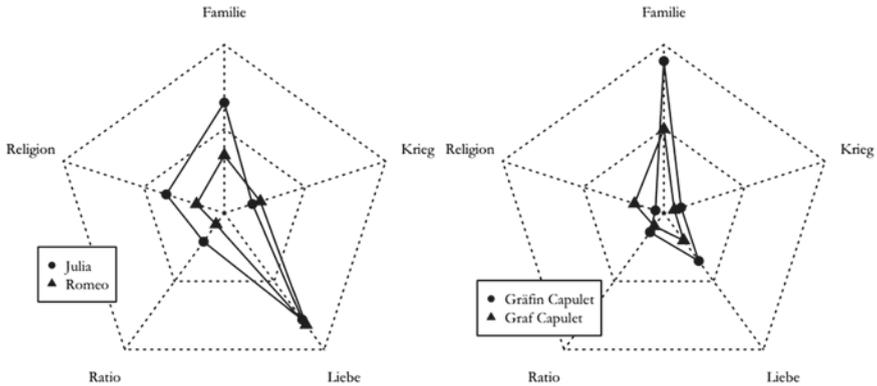


Fig. 1 Romeo's and Juliet's speech (left) and Juliet's parent's speech (right) analyzed using a custom dictionary

Enriched Custom Dictionary As second variant of the custom dictionary approach, we use the previously built custom dictionary as a set of seed words and add words that are distributionally similar. This similarity is measured as distance over word vectors created with the tool *word2vec*,¹⁴ integrated as an R package *wordVectors*.¹⁵ The word vectors have been trained on a corpus of German literary fiction.¹⁶ Although a different literary genre (prose), we selected this corpus for its size and because it covers the historic period in question. The enriched dictionaries contain the same words as before, but additionally all words that have a (cosine) similarity of 0.5 or higher. This results in between 136 and 212 entries per keyword. In total this dictionary contains 893 entries.

GermaNet Dictionary As a third, technically similar option, we test the use of *GermaNet* 11¹⁷ as a dictionary source. *GermaNet* is a concept network, primarily used for (computational) linguistic purposes (and developed by computational

¹⁴Tomas Mikolov/Ilya Sutskever/Kai Chen et al., "Distributed Representations of Words and Phrases and their Compositionality," in: *Advances in Neural Information Processing Systems* 26 (2013), 3111–3119.

¹⁵Benjamin Schmidt, *WordVectors* (2017), <https://github.com/bmschmidt/wordVectors> (retrieved on June 10, 2018).

¹⁶143M tokens in 2,737 texts; Frank Fischer/Jannik Strötgen, "When Does German Literature Take Place? – On the Analysis of Temporal Expressions in Large Corpora," in: *Proceedings of the Annual Conference of the Alliance of Digital Humanities Organizations* (2015), <https://people.mpi-inf.mpg.de/~jstroetge/papers/2015-DH-FischerStruetgen-WhenDoesGermanLiteratureTakePlace.pdf> (retrieved on June 10, 2018).

¹⁷Birgit Hamp/Helmut Feldweg, "GermaNet-a Lexical-Semantic Net for German," in: *Proceedings of the Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications on the Conference of the Association of Computational Linguistics* 1997, <https://www.aclweb.org/anthology/W97-0802> (retrieved on June 10, 2018).

linguists). A concept in *GermaNet* may or may not be lexicalized: Systematic gaps in the hierarchy have been filled with concepts, even if no German word exists for this concept. The relations between concepts are lexical relations like hypernymy/hyponymy or various forms of meronymy. The hypernymy/hyponymy-relations can be used to form a concept tree based on inheritance. While there is a technical top node in this tree, the concepts directly below this node are grouped into so-called semantic fields (see *GermaNet* Section in the Appendix).

Due to the concept hierarchy, one could in principle operate on different granularity levels. As the other systems are using a limited (and rather small) number of dimensions, we chose to use the semantic field classification, and extracted all realized words that belong to these fields (from anywhere in the hierarchy). This results in a dictionary consisting of 54 keywords, each of which containing 85 to 18,070 lemmas. In total this dictionary contains 142,815 entries.

3.2 Topic Modeling

Topic modeling has been introduced by Blei et al.¹⁸ The core idea is to model topics as distributions over words (i.e., a word has a certain probability to be in a topic), and to model documents as distributions over topics. A document is then a combination of various topics, each assigned a probability. When training a topic model, one needs to fix the number of topics the model should consider.

Within digital humanities, topic models are widely used as a means to discover thematic structure of text corpora. Underwood/Goldstone use topic models to study the history of the *Publications of the Modern Language Association* (a scientific journal);¹⁹ Rhody studies figurative language in poetry,²⁰ Quamen/Hjartarson investigate the personal correspondence of two Canadian writers,²¹ and Schöch studies dramatic genre (and sub-genre) using topic models²² (this list is not exhaustive).

For our experiments, we employ topic models in two variants, each with a low and high number of topics (20 vs. 100 topics). In the first variant, we train the

¹⁸David Blei/Andrew Y. Ng/Michael I. Jordan, “Latent Dirichlet Allocation,” in: *Journal of Machine Learning Research* 3 (2003), 993–1022.

¹⁹Ted Underwood/Andrew Goldstone, “What can topic models of PMLA teach us about the history of literary scholarship?” in: *Journal of Digital Humanities* 2/1 (2014), <http://journalofdigitalhumanities.org/2-1/what-can-topic-models-of-pmla-teach-us-by-ted-underwood-and-andrew-goldstone/> (retrieved on June 10, 2018).

²⁰Lisa M. Rhody, “Topic Modeling and Figurative Language,” in: *Journal of Digital Humanities* 12/1 (2012), 19–38.

²¹Harvey Quamen/Paul Hjartarson, “Big Data and the Literary Archive: Topic Modeling the Watson-McLuhan Correspondence,” in: *Proceedings of Dh* (2014), <http://dharchive.org/paper/DH2014/Poster-94.xml> (retrieved on June 10, 2018).

²²Schöch (Ann. 6).

topic models on full text level: Each dramatic text is taken as a single document, containing the voices of all characters as one (stage directions and speaker names are removed). In the second variant, we consider the speech of every character as a new document. The entire speech of Emilia from *Emilia Galotti*, for instance, is put together in a new document, containing only her speech.

4 Postulates about Dramatic Texts

In this section, we substantiate our understanding of what we call ‘postulates about dramatic texts’, focusing on their motivation. Their operationalization will be discussed in Sect. 6.

4.1 Copresence

Following from the intuition that conversations on stage share a common theme (or set of themes), we expect that two copresent characters (which are characters with shared stage presence) share themes in their speech. This has never been subject to drama-historical postulates or poetological claims but seems plausible for our setting. In addition, it is in line with previous observations we made in former work on the translation of Shakespeare’s *Romeo and Juliet*.²³ Using the custom dictionary-approach to compare Shakespeare’s characters with those from German plays, we found that in dramatic genres with small ensembles—like the *Bürgerliche Trauerspiel*—the character speech of the most dominant characters is similar in terms of thematic distribution. That seems to be a result of ensemble size. The smaller it is, the greater the probability that characters spend time together on stage. We also found that copresence and themes of speech seems to be linked in single pairs of characters. Striking examples are in this case butlers and servants, which are often peripheral or satellite characters: Their only network connection is the one to their lord.²⁴ In the examples we give, this connection is usually not the only one, but it is the strongest. E.g., in Schillers *Kabale und Liebe*, the distribution of themes in the speech of Präsident von Walter and his butler is very similar. The same applies for Julieta and the nurse in Shakespeare’s *Romeo and Juliet* and even for Hettore Gonzaga and Marinelli in *Emilia Galotti*.

4.2 Scene Boundaries

The postulate about scene boundaries leads back to poetological controversies about the function of scenes that started with the classical unities of action, time

²³Willand/Reiter (Ann. 12).

²⁴Franco Moretti, *Distant Reading*, London 2013, 227.

and place. The latter was not introduced by Aristotle, but by J. C. Scaliger and adopted by the French *doctrine classique* in the 16th and 17th century. The neo-Aristotelian ideal of what Volker Klotz calls “geschlossenes Drama”²⁵ postulates a causal development of plot and the *liaison des scènes*.²⁶ This means that two adjacent scenes are connected by at least one character that remains present. This is supposed to ensure the *unity of action*. Gottsched—an assertor of the French tradition—describes it as follows:

Schließlich muß ich erinnern, daß die Auftritte oder Scenen in einer Handlung [meint “Akt”] allezeit mit einander verbunden seyn müssen: damit die Bühne nicht eher gantz ledig werde, bis die gantze Handlung aus ist. Es muß also aus der vorigen Scene immer eine Person da bleiben, wenn eine neue kommt, oder eine abgeht: damit die gantze Handlung einen Zusammenhang habe. Die Alten sowohl, als Corneille und Racine haben dieses fleißig beobachtet: aber in unsern deutschen Stücken geht es gemeinlich sehr bunt durcheinander; und daher kommt es grosentheils, daß die Fabeln nicht recht zusammen hängen.

[Finally I must note that the scenes in an act must always be connected with each other: so that the stage does not become completely empty until the whole action is over. There must always be one person from the previous scene there when a new one comes, or one leaves: so that the whole act has a connection. The old ones as well as Corneille and Racine have observed this diligently: but in our German plays it is generally very colorfully mixed up; and that’s why it comes largely that the fables don’t really hang together.]²⁷

Even though Gottsched hints at the contemporarily widespread touring companies—that he wanted to be replaced by institutionalized theaters –, he also foresees a drama-historical development that would start two decades later. It basically breaks with the French *doctrine classique* and sets Shakespeare as the new poetological landmark. This development—which peaks in *Sturm und Drang* poetics—leads to our postulate about scene boundaries. It combines the structural unit ‘scene’ with the themes of character speech in a scene. If many characters remain on stage, the themes prevalent in the character speech remain similar, so that as Gottsched says, “die Fabeln [...] zusammen hängen [the threads are connected].” In contrast, if there is a major exchange of characters between the scenes, the themes spoken in those scenes are likely to change as well.

²⁵Volker Klotz, *Geschlossene und offene Form im Drama*, München 1960.

²⁶Scherer dates the *liaison des scènes* as a common rule back to 1640ies France. Cf. Jacques Scherer, *La dramaturgie classique en France*, Paris 1950. The rise and fall of the interrelations between ceremonial and dramatic principles of court representation describes Juliane Vogel, “Aus dem Takt: Auftrittstrukturen in Schillers *Don Carlos*,” in: *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* 86/4 (2012), 532–546.

²⁷Johann C. Gottsched, *Versuch einer Critischen Dichtkunst vor die Deutschen*, Leipzig 1730, 585. Transl. by the authors, NR/MW.

4.3 Tragic or Happy Ends: Die or Marry

In her 2005 Novel *Tenor of Love*, the Canadian feminist poet, author and university professor Mary di Michele writes: “Somebody always has to die onstage, die or marry; that’s the only difference between a comedy and a tragedy as far as the world knows.”²⁸ In fact, there are a lot of differences between tragedies and comedies and most of them—just as the function of scenes—are object to manifold controversies for more than 2,000 years now. Some of them are plot-related and point to the divergent endings, other differences are bound to the dramatic characters. Their historicity, moral qualities, social status and decorum of speech are the best known. Among others Scaliger, Opitz, Gottsched and Lessing discussed the question, whether middle-class characters have tragic qualities or solely higher ranked characters as kings do.²⁹

This postulate brings into focus the ending of the drama. The expectation here is that the sketched differences in events (death or marriage) are reflected in different themes appearing in the character speech. Commenting or describing someone’s wedding implies the appearance of wedding-related themes, while the death of a character can be expected to be discussed in different themes.

4.4 Dynamic and Static Protagonists

Characters in drama can be considered as either dynamic or static.³⁰ In antique tragedy, e.g., actors wore masks (*persona*) that expressed a fixed, stable character property, like a specific emotion. This mask represented the character’s set of values and the way they would deal with what fate intended for him.³¹ A change of mind would neither correspond to the anthropological model of the Greek nor to their poetological understanding of dealing with the dramatic conflict. During the course of drama history, the idea of a static character got more and more replaced by the attempt to form characters in line with modern individuals. For German tragedy, this individualization of characters is usually stated to become a broader phenomenon with the rise of the *Bürgerliche Trauerspiel* in the second half of the

²⁸ Mary Di Michele, *Tenor of Love*, New York 2005, 20.

²⁹ Cf. for a brief sketch of the so called estates-clause Harald Weinrich, “Hoch und Niedrig in der Literatur,” in: Horst Rüdiger (ed.), *Literatur und Dichtung*, Stuttgart u. a. 1973, 160–170.

³⁰ Manfred Pfister, *The Theory and Analysis of Drama*, Cambridge, UK, New York 1988, 241–243.

³¹ An outstanding description of drama-historical changes and their dependencies from the social and religious values of a time can be found in Hebbels poetological foreword to “*Maria Magdalena*, betreffend das Verhältnis der dramatischen Kunst zur Zeit und verwandte Punkte,” cf. Friedrich Hebbel, “Vorwort zur ‘Maria Magdalena’, betreffend das Verhältnis der dramatischen Kunst zur Zeit und verwandte Punkte,” in: *Sämtliche Werke*, ed. Richard M. Werner, Berlin 1911, 39–65.

18th century:³² Characters react to the dramatic conflict by changing their mind and values.

This postulate aims at the protagonist of the play and the idea that the thematic design of a dynamic character's speech is more diverse than the speech of a static character. Since our corpus covers one century starting 1730, we expect it to contain mostly dynamic protagonists and aim for automatically determining their dynamicism.

5 Experimental Setup

5.1 Formalization

Each theme detection method produces a vector for a given input text. For the dictionary-based methods, this is the (normalized) number of words in the respective set. The methods using topic modeling return a vector of probabilities with one element for each topic. We make no assumptions on interpretability of the theme vectors, nor do we inspect the topics manually. Comparing system predictions on texts or text segments typically involves a notion of distance between the vectors. For all experiments, we employ *Euclidean* distance to measure distance between vectors and *Pearson correlation*³³ to measure correlations.³⁴

A key component in the formalization of the postulates is the stage presence matrix.³⁵ This binary matrix represents scenes in columns and characters in rows. If a character has stage presence during a scene, the cell is marked with a 1 (one). All other cells contain 0 (zero). In our analysis, we equate 'being present on stage' with 'speaking on stage'. This is a simplification, as mere presence on stage is not encoded in a machine-readable fashion in our corpus. Interestingly, this distinction can already be found in Gryphius's plays, e.g., in *Carolus Stuardus*, written about 1650: Not all characters present in a scene do actually speak, and so Gryphius differentiates in the *Dramatis Personae* of his plays between characters that enter stage 'as speakers' and those that enter 'as mute' characters.³⁶

³² Cf. Bernhard Asmuth, *Einführung in die Dramenanalyse*, Stuttgart ⁵1997.

³³ Karl Pearson, "Notes on Regression and Inheritance in the Case of Two Parents," in: *Proceedings of the Royal Society of London* 58 (1895), 240–242.

³⁴ *Ibid.* As we compare different results, it is important to keep this setting constant over all our experiments. Other choices have not been explored.

³⁵ This is of course related to the configuration matrix (cf. Pfister [Ann. 30]), but based on scenes instead of configurations.

³⁶ Cf. Albrecht C. Rotth (*Vollständige Deutsche Poesie: In drey Theilen*, Leipzig 1688, 141), who postulates: "Die Zahl der Scenen in welche man die Actus wiederum eintheilet/stehen dem Poeten frey. Doch nimmt man ordentlicher Weise über 5. oder 6. derselben nicht gerne/führt auch über 4. Personen redend nicht ein/Damit sie den Zuschauer nicht verwirren; Stumme Personen aber mögen seyn so viel ihr woollen."

$$M = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

To give an example, we assume the stage presence matrix shown in Eq. 1. Character 1 is therefore present on stage in scenes one and three. Individual columns or rows can be extracted: The vector for Scene 2 is $[0, 1, 1, 0] = M[:, 2]$ (second column), and the vector for character 3 is $[0, 1, 0] = M[3, :]$ (third row).

$$C_i = \{x | M[x, i] = 1\} \quad (2)$$

Based on this matrix, we can define the set of characters in a scene (C_i) as shown in Eq. 2: All characters that have a 1 in the respective column. C will be the set of all characters in the play. We will use both the matrix and set representation for operationalizing the postulates.

5.2 Corpus

All texts are taken from the *TextGrid repository*³⁷ and have been written or translated into German between 1730 and 1930. We selected texts that can be classified (mostly) uncontroversially as either being a tragedy or a comedy. A table that lists the entire corpus can be found in the *Appendix*.

5.3 Preprocessing

The texts and their structure have been converted into stand-off annotations in the UIMA pipeline framework.³⁸ Stage directions and character speech have been processed separately, to ensure that inserted stage directions do not break the syntactic structure of the surrounding sentences. This has been achieved with the *DramaNLP* package.³⁹ The texts have been lemmatized and part of speech-tagged, using the Stanford part of speech-tagger⁴⁰ and the Mate lemmatizer⁴¹,

³⁷ <https://textgridrep.org> (retrieved on June 10, 2018).

³⁸ <http://uima.apache.org> (retrieved on June 10, 2018).

³⁹ Nils Reiter, *DramaNLP* (2017), [10.5281/zenodo.400627](https://zenodo.org/record/400627).

⁴⁰ Kristina Toutanova/Dan Klein/Christopher Manning et al., “Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network,” in: *Proceedings of HLT-NAACL* (2003), 252–259.

⁴¹ Anders Björkelund/Bernd Bohnet/Love Hafdell et al., “A High-Performance Syntactic and Semantic Dependency Parser,” in: *Coling 2010: Demonstrations*, 33–36, <http://www.aclweb.org/anthology/C10-3009> (retrieved on June 10, 2018).

integrated into the *DKpro core* package⁴². Speaker attributions have been made using manually specified mappings in difficult cases and string overlap heuristics in other cases.

After preprocessing, the texts have been exported into CSV files and are analyzed with *RStudio*, using the R package *DramaAnalysis*.⁴³

6 Experiments

In this section, we will describe in detail the operationalization of the postulates introduced above and the results, separately for each postulate.

6.1 Copresence

Postulate: Copresence of characters leads to similar thematic profiles of their speech.

Operationalization

Given the scene presence matrix M for a play, we first create a copresence matrix by multiplying the presence matrix with its transpose: $M_c = MM^T$. This (quadratic, symmetric) matrix contains the characters as rows and columns, and the number of scenes in which two characters are co-present as cell values. As a next step, this matrix is converted to a distance matrix by inverting and normalizing with the total number of scenes (s , equivalent to the number of columns in the stage presence matrix): $M_{dist} = 1 - (M_c/s)$.

This distance matrix contains values between 0 and 1. Pairs of characters that appear in more scenes together have lower values (lower distance) than characters that appear rarely together. Character pairs that are not co-present at all have the maximal distance of 1, character pairs that are co-present in all scenes have a distance of 0. Characters that speak less than 1,000 words are excluded from this analysis. As an example, Table 2 shows the distance matrix for *Emilia Galotti*. Emilia and Orsina do not share a single scene, which is marked by the maximal value in the corresponding cell. Emilia and Appiani share one single scene, which gives them a distance score of 0.98.

We calculate the correlation between distances based on co-presence and *Euclidean* distances between theme vectors. A positive correlation indicates that

⁴²Richard Eckart de Castilho/Iryna Gurevych, “A broad-coverage collection of portable NLP components for building shareable analysis pipelines,” in: *Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT*, Dublin (2014), 1–11, <http://www.aclweb.org/anthology/W14-5201> (retrieved on June 10, 2018).

⁴³Version 2, Nils Reiter, *DramaAnalysis*, 2017, [10.5281/zenodo.847167](https://doi.org/10.5281/zenodo.847167).

Table 2 Distance matrix for Emilia Galotti

	Appiani	Claudia G	Der Prinz	Emilia	Marinelli	Odoardo G
Claudia G	0.91					
Der Prinz	1.00	1.00				
Emilia	0.98	0.93	0.95			
Marinelli	0.95	0.93	0.79	0.91		
Odoardo G	1.00	0.93	0.95	0.95	0.91	
Orsina	1.00	0.98	0.98	1.00	0.93	0.93

co-presence and shared topics in character speech go hand in hand, a negative correlation indicates the opposite. A value of zero indicates no correlation between the two. The correlation metric reflects how strong this postulate is realized in texts, and there is no way of formulating an exact expectation. High correlations are unrealistic to expect, as exceptions to the postulate come to mind relatively quick. Our discussion focuses on the range between 0.1 and 0.5.

Results

Figure 2 shows the percentage of dramas that fulfill the copresence postulate, for various correlation thresholds (on the x-axis). For instance, just over 75 % of the plays have a positive correlation (strictly > 0) when using the GermaNet dictionary method.

In general, we see similar curves for all methods, although the difference between the best and worst method is up to 10–15 percentage points (for low correlations). For correlations below 0.5, dictionary methods achieve the better performances. For higher correlations, the custom dictionary method becomes one of the worst, while enriched and *GermaNet* are still achieving top performance rates. The four topic modeling variants are outperformed in most settings. It is a surprising result that methods that cover only a fixed set of themes (custom and enriched dictionary) do perform on a similar level as methods that are in principle covering all themes (*GermaNet* dictionary and topic modeling).

The fact that all methods achieve low correlation scores across the board suggests that this concept (of some copresence leading to similar themes) is hard to grasp for automatic methods. One reason is that characters are usually copresent with more than one other character throughout the play—the entire text a character speaks is actually a mixture of contributions to thematically more or less connected situations (scenes), in which the content is also influenced by other present characters. Another difficulty is that copresence is not limited to two present characters—it's not unusual that three or more characters interact on stage simultaneously. This polyphony of character speeches influences the speech of each character. None of the methods is designed to cope with these problems.

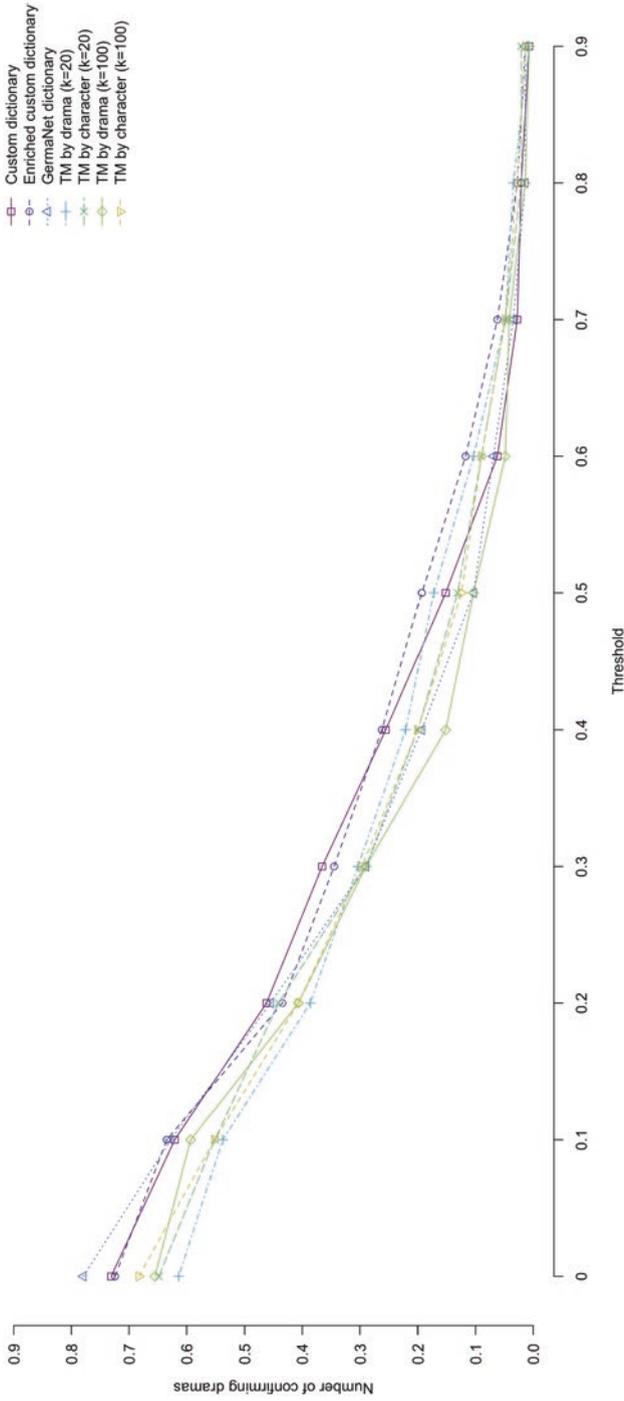


Fig. 2 Percentage of plays that fulfill the copresence postulate, given the correlation threshold and theme detection method

6.2 Scene boundaries

Postulate: The more characters are exchanged at scene boundaries, the more different the two scenes are thematically.

Operationalization

To our knowledge, two possible ways of quantifying the change of stage personnel have been published: Marcus proposes the use of the *Hamming distance* on a binary stage presence matrix.⁴⁴ Given two symbol sequences of equal length, the *Hamming distance* is the number of positions in which the two sequences are different. Applied to two scene vectors extracted from a scene presence matrix, this is the number of characters that leave or enter the stage at a scene boundary.

Hamming distance can be scaled to the interval between zero and one in two ways: Dividing by the total number of characters ($|C|$) allows comparison across dramatic texts, as a play with a large dramatis personae can exchange more characters than one with a small cast (Eq. 3). Trilcke et al. propose the *drama change rate* as a new measure to quantify this change.⁴⁵ *Drama change rate* is almost equivalent to the *Hamming distance*, but it uses the number of characters in the two adjacent scenes to scale (Eq. 4), not the entire cast. In all variants, the measure is minimal if all characters remain on stage. Scaled *Hamming distance* is only maximal if all characters in the play are affected, e.g., by one half leaving and the other half entering the stage. *Drama change rate* is maximal if all characters that are present in one of the two scenes are exchanged, irrespective of their proportion at the full dramatis personae.

$$H_i = \frac{|C_i \Delta C_{i+1}|}{|C|} \quad (3)$$

$$r_i = \frac{|C_i \Delta C_{i+1}|}{|C_i \cup C_{i+1}|} \quad (4)$$

Draxler describes the *scenic difference* (*szenische Differenz*) between two neighboring scenes as the total number of characters in the drama minus the number of characters present in both scenes (Eq. 5).⁴⁶ The resulting number is maximal (i.e., equal to the number of characters in the play), if not a single

⁴⁴Solomon Marcus, *Mathematische Poetik*, Bucuresti 1973; Richard W. Hamming, “Error detecting and error correcting codes,” in: *The Bell System Technical Journal* 29/2 (1950), 147–160, [10.1002/j.1538-7305.1950.tb00463.x](https://doi.org/10.1002/j.1538-7305.1950.tb00463.x) (retrieved on June 10, 2018).

⁴⁵Peer Trilcke/Frank Fischer/Mathias Göbel et al., “Netzwerkdynamik und Plotanalyse. Zur Visualisierung und Berechnung der ‘progressiven Strukturierung’ literarischer Texte,” in: *Book of Abstracts of Dhd 2017*, Bern 2017, 175–180.

⁴⁶Christoph Draxler, “Computerunterstützte Dramenanalyse,” Master’s thesis, LMU Munich 1988.

character is present in both scenes (the intersection becomes zero). It is minimal if *all* characters in the entire play remain on stage.⁴⁷ The *scenic difference* can be scaled by dividing by the total number of characters in the play.

$$s_i = |C| - |C_i \cap C_{i+1}| \quad (5)$$

The difference between the measures can first be characterized as one of scope: Scaled *Hamming distance* directly expresses how many percent of the dramatis personae are exchanged at this boundary. *Drama change rate* is a local measure, ignoring whether the change concerns 90 or 10 percent of the dramatis personae. *Scenic difference* gives a number that is relative to the total number of characters in the dramatis personae. Only if all characters in the play are present in either one of the scenes ($C = C_1 \cup C_2$) or the number of remaining characters is zero ($C_1 \cap C_2 = \emptyset$), the measures produce the same number. The difference between the measures grows with the number of characters in the dramatis personae, and with the stability of the stage personnel.

It is noteworthy that Draxler's *scenic difference* sometimes produces counter-intuitive results: Consider two scene boundaries as an example. A single character is entering the stage at the first, while a group of characters is entering at the second. The intersection of characters across the boundaries will remain the same at both boundaries, leading to the same *scenic difference*—although the group entering at the second boundary might be as large as the entire set of characters already on stage.

Figure 3 shows plotted *Hamming distance/change rate/scenic difference* (dashed lines) and thematic distance values (solid line) for the Lessing play *Emilia Galotti* and the Schiller play *Jungfrau von Orleans*. The vertical lines depict act boundaries. Broadly speaking, the measures make similar movements at most scene boundaries. Act boundaries show higher rates. *Scenic difference* and *Hamming distance* values are generally in a smaller interval, as they express change rates relative to the entire set of characters.

The difference between *Hamming distance* and *drama change rate* (that results from different scaling methods) becomes apparent between scenes 17–20 in *Die Jungfrau von Orleans* (bottom plot). *Drama change rate* shows a continuous maximal line (as all characters are exchanged at the boundaries), while the scaled *Hamming distance* differentiates between 15 and 20 % of the characters being exchanged.

According to the postulate, we expect a correlation between the thematic distance (solid line) and one of the change rates (dashed lines). We do *not* expect a perfect correlation (this becomes obvious in the examples). In some cases, the same set of themes is continued across scene boundaries (at least according to this

⁴⁷ Ibid. correctly notes that this case cannot occur. His analysis is based on configurations, and configuration boundaries are defined by the fact that a character enters or leaves the stage. Our analysis is based on scene boundaries, and it could well be the case that no personnel change takes place.

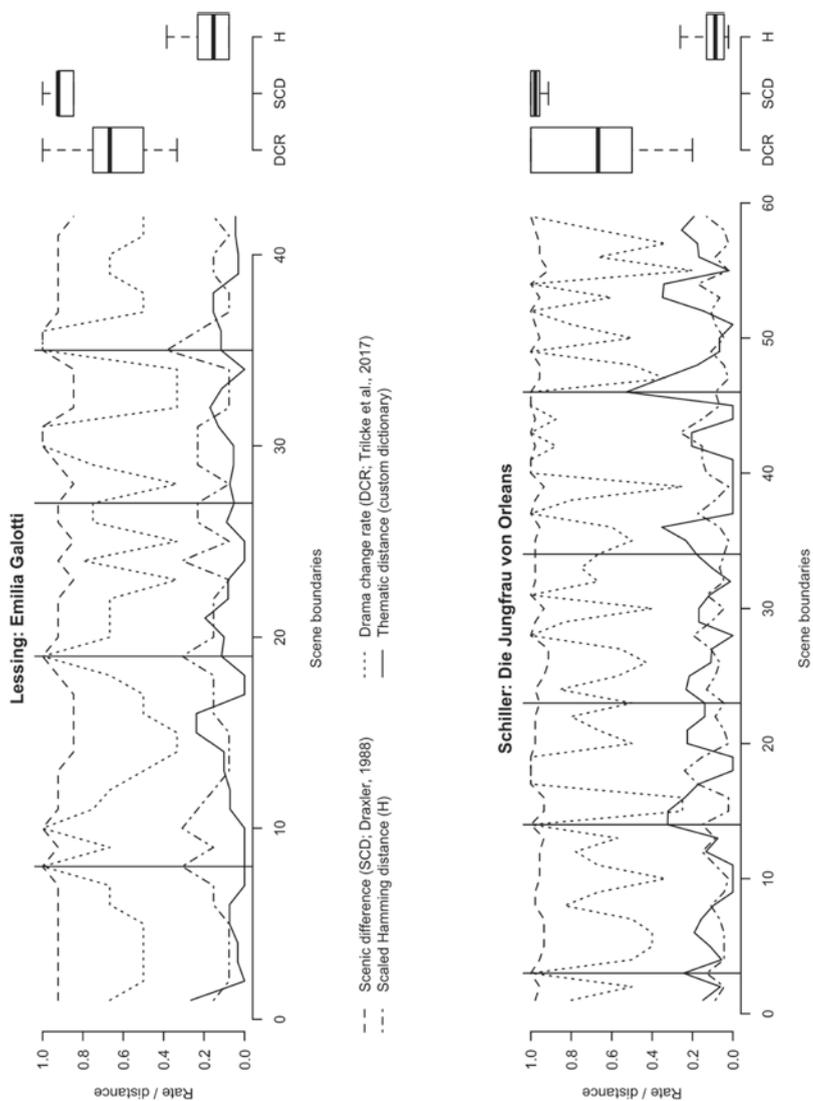


Fig. 3 Drama change rate, scenic differences and distances calculated based on a custom dictionary for two plays. Vertical lines represent act or prolog boundaries

theme detection method) even if characters are exchanged (e.g., across the first act boundary in *Emilia Galotti*). In some parts, however, we see a similar movement (e.g., the peak in the last act of the *Jungfrau von Orleans*).

Technically, we calculate *Pearson correlations* between one of the measures at each scene boundary, and the *Euclidean* distance between the scene before and after the boundary. Scene boundaries that are also act boundaries are *not* removed, as we expect them to behave in line with the postulate.

Results

Figure 4 presents the results in a similar way as for the previous postulate. Again, the x-axis shows the correlation threshold and the y-axis the number of texts for which we find a positive correlation according to the postulate. The left plot shows the positive correlations with the *drama change rate*, the right plot the positive correlations with *scenic difference*.

The most striking observation here is that the results are better differentiated when correlated with the *scenic difference*. In terms of method ranking, however, the results are similar: The custom dictionary method clearly performs worst, with scores over 10 pp. behind the system ranked second last. To a great extent, this is probably related to the limited size of only five discriminatory topics. That they might be relevant for a lot of plays does not implicate, that they are the *only* themes discussed in plays. Themes like death or nature are simply not recognised by the custom dictionaries. Enriched dictionary performs a bit better, but is still outperformed by the others, at least up to a correlation coefficient of 0.4. The different variants of topic modeling perform all similar—depending on the correlation level, one or the other variant is slightly in the lead. A varying number of dimensions does not make a big difference, although topic models with more dimensions are slightly in the lead.

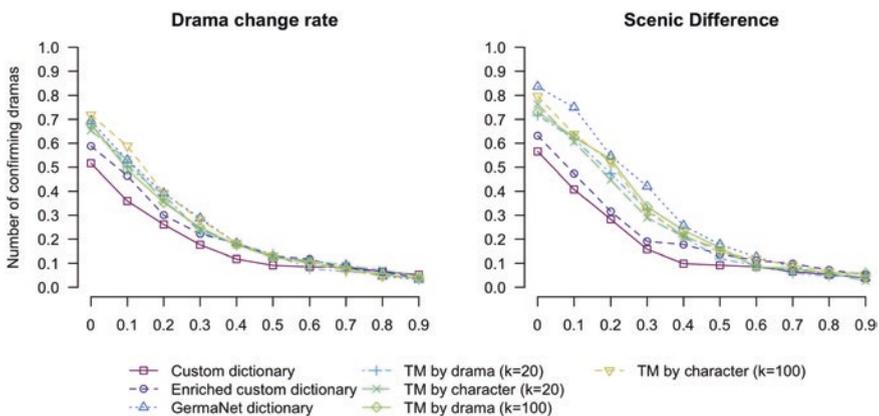


Fig. 4 Percentage of plays that fulfill the scene boundary postulate, given the correlation threshold and theme detection method

The *GermaNet* dictionary method performs best if correlated against the *scenic difference*. One possible explanation could be that the ‘counter-intuitive’ scores this measure sometimes produces (see above) are actually in line with the postulate. Characters entering the stage might not make a big difference in terms of discussed themes.

A further possible reason for generally low correlation scores might be the diverging understandings of the function of scenes in drama history. While boundaries between *acts* in almost all plays mark a change of time or place, boundaries between *scenes* are known to be ambiguous.⁴⁸ That is mainly because the sketched French and German post-Gottsched tradition in poetics and playwriting differ in terms and concepts. In the French tradition coming from Scaliger, a play is structured by *scène* and *acte*. The German tradition since 1750—which leads back to Shakespeare—distinguishes *Auftritt*, *Szene* and *Akt*.⁴⁹ The French *acte* corresponds to the German *Szene* and *Akt* by the mentioned time/place-change.⁵⁰ In our understanding, a scene is not necessarily a plot-related category. It is less likely to be plot-related in the French *scène*, where entries (*Auftritte*) or exits of only one character cause scene boundaries. But it is more likely to be plot-related in Shakespeare’s tradition, where scene boundaries demark a change of the complete setting and the new location might imply other themes, even though the same characters are on stage.

6.3 Tragic or happy ends: Die or marry

Postulate: Tragedies and comedies are distinguishable based on the theme of the ending.

Operationalization

This postulate focuses on the ending. We therefore extract the ending from each text by considering only the last act, the last scene, or the last two scenes (determining the actual ending of a literary text is surprisingly difficult⁵¹). We consider the entire text in this ending, and disregard individual characters. Using the different methods, we generate theme vectors for each ending, and cluster

⁴⁸Cf. Bernhard Asmuth, “Charakter;” in: Klaus Weimar/Harald Fricke/Klaus Grubmüller et al. (eds.), *Reallexikon der deutschen Literaturwissenschaft*, Berlin 1997, 297–299, here: 37–40.

⁴⁹Pfister (Ann. 30), 238.

⁵⁰The English term ‘scene’ is used as diverse as the German or French; cf. Bruce Smith, “Scene;” in: Henry S. Turner (ed.), *Early modern theatricality*, Oxford 2013, 93–112.

⁵¹Cf. Fotis Jannidis/Isabella Reger/Albin Zehe et al., “Analyzing Features for the Detection of Happy Endings in German Novels,” in: *DHd2017: Book of Abstracts*, Bern 2017, 81–85, arXiv:1611.09028 [cs.IR] (November 28, 2016), <https://arxiv.org/abs/1611.09028>, http://www.dhd2017.ch/wp-content/uploads/2017/03/Abstractband_def3_M%C3%A4rz.pdf (retrieved on June 10, 2018).

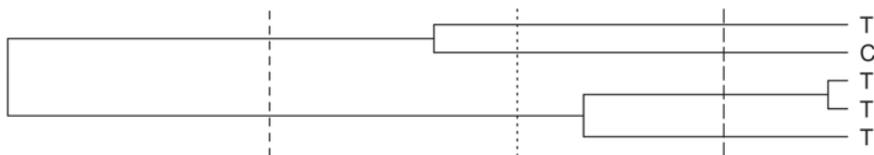


Fig. 5 Hierarchical clustering of the endings of five plays (four tragedies, one comedy). Vertical lines show groupings that can be extracted

the resulting vectors hierarchically, using group-average linkage. We analyze the resulting cluster purity at multiple stages in the hierarchical clustering process, which results in differing number of clusters.

Figure 5 shows a hierarchical clustering of several plays as an example, based on the last act. Tragedies are marked with a T, comedies are marked with a C. In the bottom part, three of the four tragedies are grouped together nicely. In the top part, a comedy and a tragedy have been judged as similar. The vertical lines in the plot indicate possible stages to extract the actual groupings. Extracting the clusters at the short-dashed line yields two clusters, extracting at the dotted line yields three and the long-dashed line four clusters.

$$P(C_i) = \frac{\max_j p_i^j}{|C_i|} \quad (6)$$

Cluster purity is calculated as shown in Eq. 6, with C_i being a cluster, and the numerator the number of instances of the majority class in the cluster. In essence, cluster purity is the percentage the majority class in each cluster has. If one would use the clustering to assign class labels to each instance, cluster purity expresses the classification accuracy. Since our corpus contains more tragedies than comedies, a majority baseline would achieve a classification accuracy of 54.72% (which is equivalent with the cluster purity when putting all endings in a single cluster). In the example in Fig. 5, the clusters at the short-dashed line (left) have an average cluster purity of 0.75, the two other stages an average purity of 1.

The reasoning behind looking at multiple numbers of clusters is that both tragedy and comedy are broad, generic genres that combine many sub genres. A system might be better at distinguishing more fine-grained classes (as the one in Fig. 5), which would still be indicative of its performance for theme detection.

Results

Figure 6 shows the results for the different systems and ending variants. In each plot, we see the number of clusters on the x-axis, and the achieved cluster purity on the y-axis.

When considering the last act as the ending, two groups of systems can be discerned: The systems based on topic models, and the dictionary-based systems. The former outperform the latter clearly. For the distinction between the genres, the

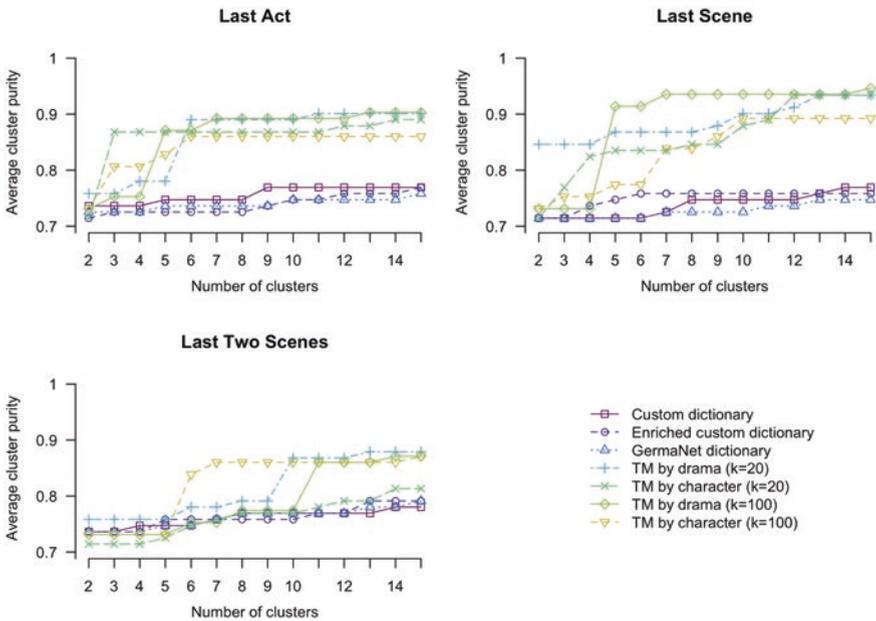


Fig. 6 Cluster purity for hierarchical clustering based on complete linkage and *Euclidean* distance

low-dimensional topic model (TM) trained on entire texts performs best. A jump in performance can be observed when distinguishing between six genres: The topic models trained on entire texts (both variants) achieve a cluster purity of about 90 %. Increasing the number of groups further has only minor influence on this score.

The picture changes slightly when considering only the last scene as the ending. The low dimensional TM trained on entire texts achieves high purity early on. Interestingly, the higher dimensional TM achieves a similar performance only for five and more genres. Character-based topic modeling is generally lower, but in a similar range.

If we look at the last two scenes understood as the ending, we observe different results. The gap between the best and the second-best system has been substantial (over ten percentage points) in the previously discussed plots. This gap does not exist in this setting. The topic models perform similarly as the dictionary-based systems. Only for more than five to seven genres, a gap between topic modeling and dictionary-based methods can be discerned.

One conclusion to draw here is that the ‘ending’ of a dramatic text is better represented by either the last scene or last act, but not multiple scenes. A possible explanation could be that the second last scene leads to more heterogeneous themes or introduces ‘non-ending material’. More material might counter-balance this effect, leading again to increased performance when considering the entire last act. But this might not be related to the dramatic ending in particular.

The additional text could be distinctive for the dramatic genre without any thematic connection to the prototypical drama endings. In terms of system choice, it is clear that topic modeling performs much better than the other methods.

6.4 Dynamic and static protagonists

Postulate: Themes in the main characters speech are changing over the course of the play.

Operationalization

To test against this postulate, we focus on the protagonist of the play. Moretti points out that identifying the actual protagonist is far from trivial—no uniform definition exist, and more complex definitions are difficult to operationalize.⁵² We therefore opt for a very simple heuristic, and select the most ‘talkative’ character, i.e., the character that speaks the most words per play.

In the three plays depicted in Fig. 7, Sara and Johanna are clearly the protagonists in Lessing’s *Miß Sara Sampson* and Schiller’s *Die Jungfrau von Orleans*—and in both plays, they are the characters with the most spoken words. There are, however, cases in which human readers do not identify the most talkative character as the main one. In Lessing’s *Emilia Galotti*, the character speaking the most is Marinelli, chamberlain of the prince. The prince himself, Emilia’s father, or Emilia, who would all be candidates for protagonists, speak, however, less than Marinelli. In line with our attempt of working with the vagueness engrained in literary works, we rely on the fact that this heuristic works in many other cases and identifies—if not the antagonist or at least one of the group of antagonists—a very *present* and therefore important character.

Given the main character in this fashion, we extract this character’s speech from the first, second, fourth and fifth act. As the key turning point in the character development is supposed to take place during Act 3, we leave out this act. Plays that do not contain five acts are excluded from this analysis.

The segmented character speech is then processed by the theme detection methods, which results in a theme vector for each act (for each main character). We then calculate *Euclidean* distance between the vectors and compare distances before/after and across the change. A distance before the change would, e.g., be the distance between first and second act, a distance across the change one between the first and fourth act. The expectation here is that for dynamic characters, the distance across the third act should be higher than before or after the third act.

⁵²Franco Moretti, “‘Operationalizing’: or, the function of measurement in modern literary theory,” in: *Pamphlets of the Stanford Literary Lab* 6 (2013), <https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf> (retrieved on June 10, 2018).

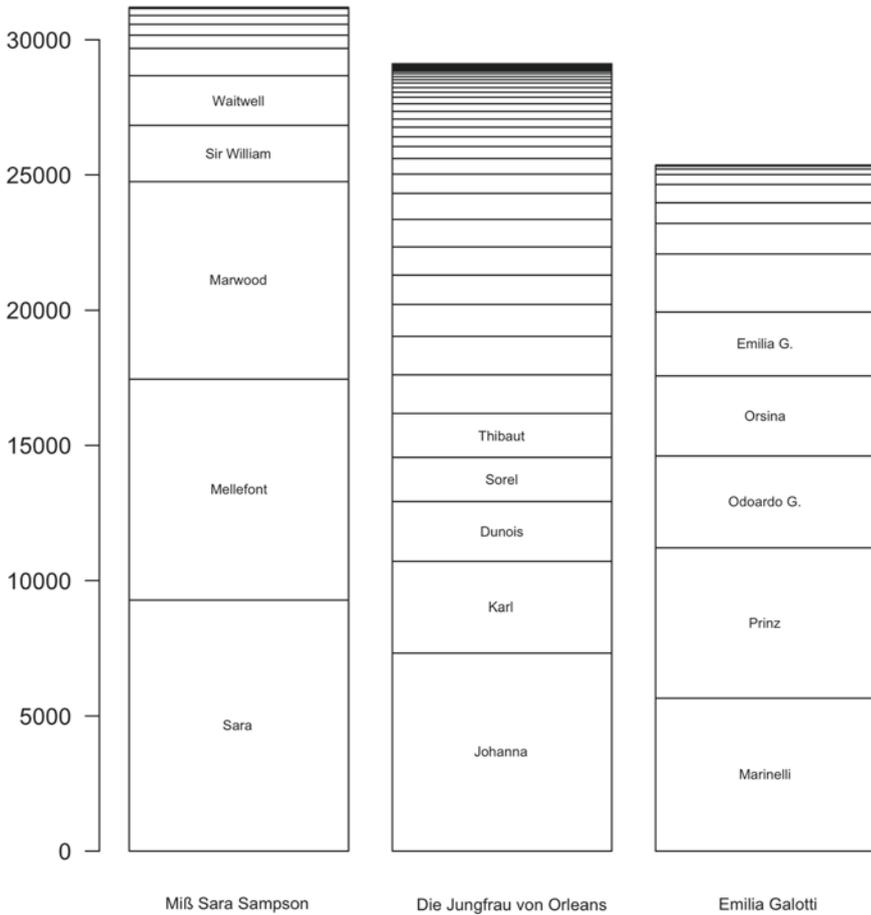


Fig. 7 Speech distribution in three plays

Results

The results of this experiment are shown in Fig. 8. Each plot shows the (normalized) distance values between one act pair from within the first or second half (x-axis) and one act pair across the third act (y-axis). Each symbol represents one drama and its act pair distances. The straight lines are linear trend lines, set in such a way that the gap between the line and the corresponding points is minimal.

A system that is able to fully represent the postulate would be shown as many points appearing in the top-left triangle, to the left or top of the diagonal. This would be caused by distances across the third act being larger than distances before/after the third act. The corresponding trend line would be relatively steep and cross the plot boundary at the top.

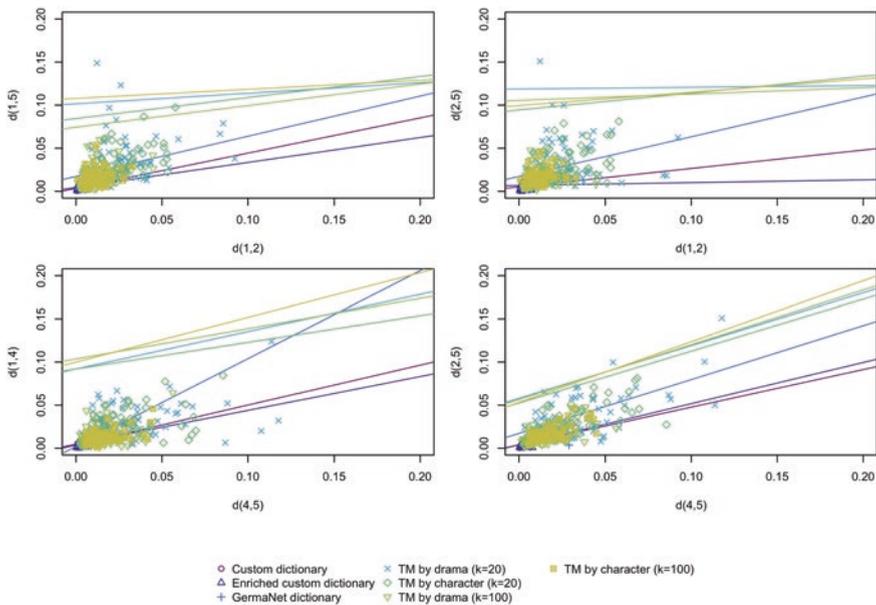


Fig. 8 Comparison of thematic distances between act pairs

It is relatively clear that this is in general not the case—the only systems with (relatively) steep trend lines are the *GermaNet*-based dictionary system and the high-dimensional character-based TM when comparing distance between act pairs 1, 4 and 4, 5 (bottom left plot). All other systems in all other settings produce gentle trend lines. In general, the methods are seemingly not able to grasp the character development appropriately.

We see two possible reasons for this bad performance: The postulated dynamics of characters might not be present on this granularity level. Developments within acts remain hidden in our analysis but could be perceived much stronger by audience or readers. This might be the case, if a character's change of mental attitude is expressed in opposing behavior: If an 'I will not do this' in the exposition becomes an 'I will do this' in the final act, theme identification methods don't stand a chance. Secondly, most of the poetological claims involving character development focus on tragedies, and it is unclear whether this postulate is supposed to hold on comedies.

7 Conclusions

This article systematically explores different ways of measuring the themes that appear in character speech in dramatic texts. The focus of this article lays on the methodology and the development of such methodology, and not

on the interpretation or use of results coming from such methods. Development and use of methodology (or tools in this case) are two distinct ways of doing computational literary studies, and it is important not to confuse the two. In order to be able to judge the quality of any method or tool, one has to have some kind of gold standard against which results are compared—even if the gold standard is not realized in a corpus or annotated data set. And if one is using a certain methodology or tool, one needs to be sure that the tool is actually doing what it is supposed to. Evaluating a tool *and* interpreting its results *at the same time* is dangerous, as it may lead to results that are not generalizable. As we have presented here, we believe it is possible to evaluate systems based on a not explicitly annotated “gold” standard.

We have compared different methods that assign themes to segments of dramatic character speech. It may not be surprising that there is no clear winner: The notion of theme is sometimes interpretative, the postulates we defined challenging, the methods not adapted to dramatic texts. We can, however, discern some strengths and weaknesses of the methods:

- Topic modeling can be applied to short text segments without severe loss of performance. Topic models that are trained on individual character’s speech perform in general similarly as models trained on entire texts, sometimes even better. We see no general difference in the scene boundary and ending postulates. Only for the copresence postulate, character-based topic models achieve less performance than models trained on the full text: This is surprising, given the fact that the text segments under investigation here *are* character speeches.
- A key difference between the *GermaNet* dictionary and the other dictionary methods is that the others cover only a fixed set of themes, while *GermaNet* is much larger and covers (in theory) most of the (modern) German language use. This difference does not lead to huge performance differences, *GermaNet* dictionary performs similar to the others in the all postulates. The only exception to this trend is the higher performance when correlated with *scenic difference* in the scene boundary postulate, but this might be more related to *scenic difference* than to the dictionary.
- Higher dimensionality of the topic models does not lead to increased performance automatically. The only case in which the higher dimensional TM performs better than its counterpart is for the third postulate on the last two scenes. In this case, the character-based TM performs better with 100 dimensions than with 20. In all other cases, the high dimensional TM is either outperformed or performs very similar to its counterpart.

The second main motivation for this paper was to experiment with a new kind of gold standard. The postulates we defined are thought to be true *in general*, which does not mean that they are true for every single text. Since many literary statements are of this kind (if made about a set of texts like a genre), a larger number of postulates can be established, which makes the entire procedure even more

fruitful. Calculating correlation scores between postulate and predicted result then incorporates the literary vagueness into the evaluation and does not hide it.

However, there are still multiple reasons to be critical about the postulates we discussed in this article: There might be disagreement on their truth, i.e., one might doubt that a postulate holds true even in general. Apart from our observations in previous work and our intuition, there is little evidence for the first postulate. The role of scene boundaries has been discussed in literary studies (cf. Sect. 4, *scene boundaries*), but there is no guarantee that playwrights adhere to what normative drama poetics tell them to. The fact that a genre ‘tragicomedy’ exists points out that the comedy/tragedy distinction is not as clear cut as it seems to be. While this might sound like the discussed postulates are not trustworthy, we argue that it is not necessary to believe in them fully. Inspecting different correlation levels, for instance, allows comparison to multiple, and potentially subjective, possible truths.

Canonization in literary studies might be another reason for reservations against using postulates as a gold standard. Many theories and theoretically justified predictions are based on a relatively small and non-representative selection, the literary canon. If one applies a postulate developed on the canon to non-canonical texts, it is indeed questionable whether it can be expected to be true, or even ‘truthful’. However, a closer look is necessary: Non-canonical texts might not be different from canonical texts in *all* aspects. As we have discussed in Reiter/Willand, popular plays, which are mostly not canonized, diverge from canonical plays in many structural aspects.⁵³ The corpus used in this work, however, consists mostly of canonical plays.

A second possible critique towards the postulates is that they might not be related to themes. We initially believed that the comparisons discussed above include a thematic change, at least to some extent. The results partly suggest that this is might not always be the case; surprisingly, the analysis of character development is such a case. While the exact extent is an open question to all our method testing scenarios, we can still not decide whether the postulate or the operationalization by theme is the crux of the matter. As an outlook for the future, it would be helpful to apply our methods to synthetic texts. These could—to stay with our example—contain thematically ideal dynamic and static characters and thus contribute to the answer to our questions.

Finally, all postulates need to be operationalized before conducting any kind of experiment. Operationalizing a concept always entails decisions that may lead to a misrepresentation of the concept. Some choices might not even be foreseen conceptually, as is best exemplified by the different quantification methods for personnel change for the scene boundary postulate. We believe that the above described operationalizations are reasonable, without saying that they are without alternatives. However, we also believe a scientific discussion about appropriate ways

⁵³Nils Reiter/Marcus Willand, “Surveying Shakespeare’s Impact on the German Drama: Taking a Computational Approach to an Epoch,” in: Sandro Jung and Michael Wood (eds.), *Anglo-German Dramatic and Poetic Cultures: New Perspectives on Exchange in the Sattelzeit*. Bethlehem, PA, pp- 117–143 (to appear 2019).

of operationalizing literary concepts is very fruitful, and that experiments as the ones above bring to light weaknesses and strengths of such concepts. The issue of operationalization is also discussed in Reiter/Willand,⁵⁴ since we believe that progress on operationalizing literary concepts is crucial for tool and method development.

In sum, this article makes an argument for a systematic methodological development. The lack of annotated reference corpora is a challenge but can be circumvented by thoughtful and reflected use of insights derived in literary studies. Directly comparing methods to solve a fixed set of problems reveals that there is not a single winner-method: Depending on assumptions about drama structure and the exact goal, different methods need to be considered.

Acknowledgements The work described in this article has been conducted in the context of the research project *QuaDrama—Quantitative Drama Analytics*⁵⁵, generously funded by the Volkswagen Foundation.

Appendix

GermaNet Concepts

POS	Fields
Adjectives	Allgemein, Bewegung, Gefuehl, Geist, Gesellschaft, Koerper, Menge, natPhaenomen, Ort, Pertonym, Perzeption, privativ, Relation, Substanz, Verhalten, Zeit
Nouns	Artefakt, Attribut, Besitz, Form, Gefuehl, Geschehen, Gruppe, Koerper, Kognition, Kommunikation, Menge, Mensch, Motiv, Nahrung, natGegenstand, natPhaenomen, Ort, Pflanze, Relation, Substanz, Tier, Tops, Zeit
Verbs	Allgemein, Besitz, Gefuehl, Gesellschaft, Koerperfunktion, Kognition, Kommunikation, Konkurrenz, Kontakt, Lokation, natPhaenomen, Perzeption, Schoepfung, Veraenderung, Verbrauch

Corpus

Title	Author
<i>Der Wissenswurm</i>	Anzengruber, Ludwig
<i>Heimg'funden</i>	Anzengruber, Ludwig
<i>Die Kreuzelschreiber</i>	Anzengruber, Ludwig
<i>Der Rauchfangkehrer</i>	Auenbrugger, Johann Leopold von
<i>Der Postzug oder die noblen Passionen</i>	Ayrenhoff, Cornelius Hermann von
<i>Industrie und Herz</i>	Bauernfeld, Eduard von
<i>Großjährig</i>	Bauernfeld, Eduard von

⁵⁴ Ibid.

⁵⁵ <https://quadrama.github.io/>(retrieved on June 10, 2018).

Title	Author
<i>Bürgerlich und Romantisch</i>	Bauernfeld, Eduard von
<i>Die Hochzeitsreise</i>	Benedix, Julius Roderich
<i>Ein Faust der That</i>	Bleibtreu, Karl
<i>Im weißen Rößl</i>	Blumenthal, Oskar
<i>Der Bookesbeutel</i>	Borkenstein, Hinrich
<i>De rode Ünnerrock</i>	Boßdorf, Hermann
<i>Faust</i>	Braun von Braunthal, Karl Johann
<i>Der Freigeist</i>	Brawe, Joachim Wilhelm von
<i>Leonce und Lena</i>	Büchner, Georg
<i>Der Mißtrauische</i>	Cronegk, Johann Friedrich von
<i>Die Gunst des Augenblicks</i>	Devrient, Philipp Eduard
<i>Die Freier</i>	Eichendorff, Joseph von
<i>Der letzte Held von Marienburg</i>	Eichendorff, Joseph von
<i>Eid und Pflicht</i>	Engel, Johann Jakob
<i>Der Frauenmut</i>	Essig, Hermann
<i>Die Journalisten</i>	Freytag, Gustav
<i>Die zärtlichen Schwestern</i>	Gellert, Christian Fürchtegott
<i>Die Betschwester</i>	Gellert, Christian Fürchtegott
<i>Der Moloch</i>	Gerhäuser, Emil
<i>Ugolino</i>	Gerstenberg, Heinrich Wilhelm von
<i>Der Großkophta</i>	Goethe, Johann Wolfgang
<i>Faust. Der Tragödie zweiter Teil</i>	Goethe, Johann Wolfgang
<i>Clavigo</i>	Goethe, Johann Wolfgang
<i>Die natürliche Tochter</i>	Goethe, Johann Wolfgang
<i>Egmont</i>	Goethe, Johann Wolfgang
<i>Der sterbende Cato</i>	Gottsched, Johann Christoph
<i>Das Testament</i>	Gottsched, Luise Adelgunde Victorie
<i>Kaiser Heinrich der Sechste</i>	Grabbe, Christian Dietrich
<i>Kaiser Friedrich Barbarossa</i>	Grabbe, Christian Dietrich
<i>Don Juan und Faust</i>	Grabbe, Christian Dietrich
<i>Herzog Theodor von Gothland</i>	Grabbe, Christian Dietrich
<i>Des Meeres und der Liebe Wellen</i>	Grillparzer, Franz
<i>Die Jüdin von Toledo</i>	Grillparzer, Franz
<i>Ein Bruderzwist in Habsburg</i>	Grillparzer, Franz
<i>Die Ahnfrau</i>	Grillparzer, Franz
<i>König Ottokars Glück und Ende</i>	Grillparzer, Franz

Title	Author
<i>Ein treuer Diener seines Herrn</i>	Grillparzer, Franz
<i>Das Urbild des Tartüffe</i>	Gutzkow, Karl
<i>Der Furchtsame</i>	Hafner, Philipp
<i>Mägera, die fürchterliche Hexe</i>	Hafner, Philipp
<i>Hanna Jagert</i>	Hartleben, Otto Erich
<i>Rosenmontag</i>	Hartleben, Otto Erich
<i>Der Diamant</i>	Hebbel, Friedrich
<i>Der Rubin</i>	Hebbel, Friedrich
<i>Herodes und Mariamme</i>	Hebbel, Friedrich
<i>Gyges und sein Ring</i>	Hebbel, Friedrich
<i>Genoveva</i>	Hebbel, Friedrich
<i>Judith</i>	Hebbel, Friedrich
<i>Kriemhilds Rache</i>	Hebbel, Friedrich
<i>Maria Magdalene</i>	Hebbel, Friedrich
<i>Agnes Bernauer</i>	Hebbel, Friedrich
<i>Die Kinder Godunófs</i>	Heiseler, Henry von
<i>Don Juan's Ende</i>	Heyse, Paul
<i>Der Rosenkavalier</i>	Hofmannsthal, Hugo von
<i>Der Unbestechliche</i>	Hofmannsthal, Hugo von
<i>Der Schwierige</i>	Hofmannsthal, Hugo von
<i>Der Turm</i>	Hofmannsthal, Hugo von
<i>Der Tod des Empedokles</i>	Hölderlin, Friedrich
<i>Ein Trauerspiel in Berlin</i>	Holtei, Karl von
<i>Ignorabimus</i>	Holz, Arno
<i>Das Gericht von St. Petersburg</i>	Immermann, Karl
<i>Andreas Hofer, der Sandwirt von Passeyer</i>	Immermann, Karl
<i>Die Opferung</i>	Kaltneker, Hans
<i>Amphitryon</i>	Kleist, Heinrich von
<i>Der zerbrochene Krug</i>	Kleist, Heinrich von
<i>Die Familie Schroffenstein</i>	Kleist, Heinrich von
<i>Faust. Ein Trauerspiel in fünf Acten</i>	Klingemann, August
<i>Die Zwillinge</i>	Klinger, Friedrich Maximilian
<i>Das leidende Weib</i>	Klinger, Friedrich Maximilian
<i>Der Tod Adams</i>	Klopstock, Friedrich Gottlieb
<i>Die deutschen Kleinstädter</i>	Kotzebue, August von
<i>Die Indianer in England</i>	Kotzebue, August von

Title	Author
<i>Die beiden Klingsberg</i>	Kotzebue, August von
<i>Die Candidaten oder Die Mittel zu einem Amte zu gelangen</i>	Krüger, Johann Christian
<i>Die Geistlichen auf dem Lande</i>	Krüger, Johann Christian
<i>Franz von Sickingen</i>	Lassalle, Ferdinand
<i>Gottsched und Gellert</i>	Laube, Heinrich
<i>Julius von Tarent</i>	Leisewitz, Johann Anton
<i>Die Freunde machen den Philosophen</i>	Lenz, Jakob Michael Reinhold
<i>Der Hofmeister oder Vorteile der Privaterziehung</i>	Lenz, Jakob Michael Reinhold
<i>Der neue Menoza</i>	Lenz, Jakob Michael Reinhold
<i>Die Soldaten</i>	Lenz, Jakob Michael Reinhold
<i>Die alte Jungfer</i>	Lessing, Gotthold Ephraim
<i>Der Freigeist</i>	Lessing, Gotthold Ephraim
<i>Minna von Barnhelm, oder das Soldatenglück</i>	Lessing, Gotthold Ephraim
<i>Der junge Gelehrte</i>	Lessing, Gotthold Ephraim
<i>Der Misogyn</i>	Lessing, Gotthold Ephraim
<i>Die Juden</i>	Lessing, Gotthold Ephraim
<i>Samuel Henzi</i>	Lessing, Gotthold Ephraim
<i>Miß Sara Sampson</i>	Lessing, Gotthold Ephraim
<i>Emilia Galotti</i>	Lessing, Gotthold Ephraim
<i>Cleopatra</i>	Lohenstein, Daniel Casper von
<i>Epicharis</i>	Lohenstein, Daniel Casper von
<i>Ibrahim</i>	Lohenstein, Daniel Casper von
<i>Agrippina</i>	Lohenstein, Daniel Casper von
<i>Die Makkabäer</i>	Ludwig, Otto
<i>Der Erbförster</i>	Ludwig, Otto
<i>Der alte Bürger-Capitain</i>	Malß, Karl
<i>König Yngurd</i>	Müllner, Adolph
<i>Die Schuld</i>	Müllner, Adolph
<i>Die Schäferinsel</i>	Mylius, Christlob
<i>Das Schäferfest oder Die Herbstfreude</i>	Neuber, Friederike Caroline
<i>Das Liebeskonzil</i>	Panizza, Oskar
<i>Lucie Woodvil</i>	Pfeil, Johann Gottlob Benjamin
<i>Die verhängnisvolle Gabel</i>	Platen, August von
<i>Der romantische Ödipus</i>	Platen, August von
<i>Die politische Wochenstube</i>	Prutz, Robert Eduard
<i>Der Hypochondrist</i>	Quistorp, Theodor Johann

Title	Author
<i>Kritik und Antikritik</i>	Raupach, Ernst
<i>Graf Ehrenfried</i>	Reuter, Christian
<i>Die Fahnenweihe</i>	Ruederer, Josef
<i>Der fanatische Bürgermeister</i>	Scheerbart, Paul
<i>Okurirasúna</i>	Scheerbart, Paul
<i>Die Jungfrau von Orleans</i>	Schiller, Friedrich
<i>Kabale und Liebe</i>	Schiller, Friedrich
<i>Maria Stuart</i>	Schiller, Friedrich
<i>Die Verschwörung des Fiesco zu Genua</i>	Schiller, Friedrich
<i>Alarkos</i>	Schlegel, Friedrich
<i>Der geschäftige Müßiggänger</i>	Schlegel, Johann Elias
<i>Der Triumph der guten Frauen</i>	Schlegel, Johann Elias
<i>Canut</i>	Schlegel, Johann Elias
<i>Comedia des verlorenen Sons</i>	Schmeltzl, Wolfgang
<i>Professor Bernhardi</i>	Schnitzler, Arthur
<i>Liebelei</i>	Schnitzler, Arthur
<i>Der Vetter in Lissabon</i>	Schröder, Friedrich Ludwig
<i>Die Komödie der Irrungen</i>	Shakespeare, William
<i>Julius Cäsar</i>	Shakespeare, William
<i>Hamlet. Prinz von Dänemark</i>	Shakespeare, William
<i>König Richard III</i>	Shakespeare, William
<i>König Lear</i>	Shakespeare, William
<i>Julie</i>	Sturz, Helfrich Peter
<i>Der Bettler von Syrakus</i>	Sudermann, Hermann
<i>Die Lokalbahn</i>	Thoma, Ludwig
<i>Prinz Zerbino oder die Reise nach dem guten Geschmack</i>	Tieck, Ludwig
<i>Agnes Bernauerin</i>	Törring, Josef August von
<i>Ernst Herzog von Schwaben</i>	Uhland, Ludwig
<i>Faust III</i>	Vischer, Friedrich Theodor
<i>Faust. Trauerspiel mit Gesang und Tanz</i>	Voß, Julius von
<i>Die Kindermörderin</i>	Wagner, Heinrich Leopold
<i>Frühlings Erwachen</i>	Wedekind, Frank
<i>Die Büchse der Pandora</i>	Wedekind, Frank
<i>Erdgeist</i>	Wedekind, Frank
<i>Bäurischer Machiavellus</i>	Weise, Christian
<i>Atreus und Thyest</i>	Weiß, Christian Felix

Title	Author
<i>Das Manuscript</i>	Weißenthurn, Johanna von
<i>Welcher ist der Bräutigam</i>	Weißenthurn, Johanna von
<i>Lady Johanna Gray</i>	Wieland, Christoph Martin
<i>Gracchus der Volkstribun</i>	Wilbrandt, Adolf von
<i>Die Karolinger</i>	Wildenbruch, Ernst von
<i>Armut</i>	Wildgans, Anton

References

All digital references were last viewed on June 10, 2018.

- Algee-Hewitt, Mark, “Distributed Character: Quantitative Models of the English Stage, 1500–1920,” in: *Book of abstracts* (2017), 119–121, <https://dh2017.adho.org/abstracts/DH2017-abstracts.pdf>.
- Asmuth, Bernhard, “Charakter,” in: Klaus Weimar/Harald Fricke/Klaus Grubmüller et al. (eds.), *Reallexikon der deutschen Literaturwissenschaft*, Berlin 1997, 297–299.
- Asmuth, Bernhard, *Einführung in die Dramenanalyse*, Stuttgart 1997.
- Björkelund, Anders/Bohnet, Bernd/Hafdell, Love et al., “A High-Performance Syntactic and Semantic Dependency Parser,” in: *Coling 2010: Demonstrations*, 33–36, <http://www.aclweb.org/anthology/C10-3009>.
- Blei, David/Ng, Andrew Y./Jordan, Michael I., “Latent Dirichlet Allocation,” in: *Journal of Machine Learning Research* 3 (2003), 993–1022.
- Bögel, Thomas/Gertz, Michael/Gius, Evelyn et al., “Collaborative Text Annotation Meets Machine Learning: heureCLÉA, a Digital Heuristic of Narrative,” in: *DHCommons* 1 (July 2015), DOI <https://doi.org/10.5281/zenodo.3240591>.
- Bullard, Joseph/Ovesdotter Alm, Cecilia, “Computational Analysis to Explore Authors’ Depiction of Characters,” in: *Proceedings of the 3rd Workshop on Computational Linguistics for Literature* (2014), 11–16, <http://www.aclweb.org/anthology/W14-0902>.
- Draxler, Christoph, “Computerunterstützte Dramenanalyse,” Master’s thesis, LMU Munich 1988.
- Eckart de Castilho, Richard/Gurevych, Iryna, “A broad-coverage collection of portable NLP components for building shareable analysis pipelines,” in: *Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for Hlt*, Dublin (2014), 1–11, <http://www.aclweb.org/anthology/W14-5201>.
- Fischer, Frank/Strötgen, Jannik, “When Does German Literature Take Place?—On the Analysis of Temporal Expressions in Large Corpora,” in: *Proceedings of the Annual Conference of the Alliance of Digital Humanities Organizations* (2015), <https://people.mpi-inf.mpg.de/~jstroetge/papers/2015-DH-FischerStroetgen-WhenDoesGermanLiteratureTakePlace.pdf>.
- Fisseni, Bernhard/Kurji, Aadil/Löwe, Benedikt, “Annotating with Propp’s Morphology of the Folktale: Reproducibility and Trainability,” in: *Literary and Linguistic Computing* 29/4 (2014), 488–510, DOI <https://doi.org/10.1093/lc/fqu050>.
- Gottsched, Johann C., *Versuch einer Critischen Dichtkunst vor die Deutschen*, Leipzig 1730.
- Hamming, Richard W., “Error detecting and error correcting codes,” in: *The Bell System Technical Journal* 29/2 (1950), 147–160, DOI <https://doi.org/10.1002/j.1538-7305.1950.tb00463.x>.
- Hamp, Birgit/Feldweg, Helmut, “GermaNet—a Lexical-Semantic Net for German,” in: *Proceedings of the Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications on the Conference of the Association of Computational Linguistics* 1997, <https://www.aclweb.org/anthology/W97-0802>.

- Hebbel, Friedrich, “Vorwort zur ‘Maria Magdalena’, betreffend das Verhältnis der dramatischen Kunst zur Zeit und verwandte Punkte,” in: *Sämtliche Werke*, ed. Richard M. Werner, Berlin 1911, 39–65.
- Husby, Stephanie/Barbosa, Denilson, “Topic Classification of Blog Posts Using Distant Supervision,” in: *Proceedings of the Workshop on Semantic Analysis in Social Media 2012*, 28–36, <http://www.aclweb.org/anthology/W12-0604>.
- Jannidis, Fotis/Reger, Isabella/Zehe, Albin et al., “Analyzing Features for the Detection of Happy Endings in German Novels,” in: *DHd2017: Book of Abstracts*, Bern 2017, 81–85, [arXiv:1611.09028](https://arxiv.org/abs/1611.09028) [cs.IR] (November 28, 2016), <https://arxiv.org/abs/1611.09028>, http://www.dhd2017.ch/wp-content/uploads/2017/03/Abstractband_def3_M%3C%A4rz.pdf.
- Karsdorp, Folgert/Kestemont, Mike/Schöch, Christof et al., “The Love Equation: Computational Modeling of Romantic Relationships in French Classical Drama,” in: Mark A. Finlayson/Antonio Lieto/Ben Miller (eds.), *6th Workshop on Computational Models of Narrative* 45 (2015), 98–107, DOI <https://doi.org/10.4230/OASICS.CMN.2015.98>.
- Klotz, Volker, *Geschlossene und offene Form im Drama*, München 1960.
- Marcus, Solomon, *Mathematische Poetik*, Bucuresti 1973.
- Michele, Mary Di, *Tenor of Love*, New York 2005.
- Mikolov, Tomas/Sutskever, Ilya/Chen, Kai et al., “Distributed Representations of Words and Phrases and their Compositionality,” in: *Advances in Neural Information Processing Systems* 26 (2013), 3111–3119.
- Moretti, Franco, “Network Theory, Plot Analysis,” in: *Pamphlets of the Stanford Literary Lab* 2 (2011), <https://litlab.stanford.edu/LiteraryLabPamphlet2.pdf>.
- Moretti, Franco, *Distant Reading*, London 2013.
- Moretti, Franco, “‘Operationalizing’: or, the function of measurement in modern literary theory,” in: *Pamphlets of the Stanford Literary Lab* 6 (2013), <https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf>.
- Nalnick, Eric T./Baird, Henry S., “Character-to-Character Sentiment Analysis in Shakespeare’s Plays,” in: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Vol. 2: *Short Papers*, Sofia 2013, 479–483, <http://www.aclweb.org/anthology/P13-2085>.
- Pearson, Karl, “Notes on Regression and Inheritance in the Case of Two Parents,” in: *Proceedings of the Royal Society of London* 58 (1895), 240–242.
- Pfister, Manfred, *The Theory and Analysis of Drama*, Cambridge, UK, New York 1988.
- Quamen, Harvey/Hjartarson, Paul, “Big Data and the Literary Archive: Topic Modeling the Watson-McLuhan Correspondence,” in: *Proceedings of Dh* (2014), <http://dharchive.org/paper/DH2014/Poster-94.xml>.
- Reiter, Nils, *DramaAnalysis* (2017), DOI <https://doi.org/10.5281/zenodo.847167>.
- Reiter, Nils, *DramaNLP* (2017), DOI <https://doi.org/10.5281/zenodo.400627>.
- Nils Reiter/Marcus Willand, “Surveying Shakespeare’s Impact on the German Drama: Taking a Computational Approach to an Epoch,” in: Sandro Jung and Michael Wood (eds.), *Anglo-German Dramatic and Poetic Cultures: New Perspectives on Exchange in the Sattelzeit. Bethlehem, PA, pp- 117–143* (to appear 2019).
- Reiter, Nils/Gius, Evelyn/Strötgen, Jannik et al., “A Shared Task for a Shared Goal—Systematic Annotation of Literary Texts,” in: *Digital Humanities 2017: Conference Abstracts*, Montreal 2017, <https://dh2017.adho.org/abstracts/192/192.pdf>.
- Rhody, Lisa M., “Topic Modeling and Figurative Language,” in: *Journal of Digital Humanities* 12/1 (2012), 19–38.
- Roth, Albrecht C., *Vollständige Deutsche Poesie: In drey Theilen*, Leipzig 1688.
- Scherer, Jacques, *La dramaturgie classique en France*, Paris 1950.
- Schmidt, Benjamin, *WordVectors* (2017), <https://github.com/bmschmidt/wordVectors>.
- Schmidt, Benjamin, “Words Alone: Dismantling Topic Models in the Humanities,” in: *Journal of Digital Humanities* 2/1 (2012), <http://journalofdigitalhumanities.org/2-1/words-alone-by-benjamin-m-schmidt/>.
- Schöch, Christof, “Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama,” in: *Digital Humanities Quarterly* 2 (November 2017), DOI <https://doi.org/10.5281/zenodo.166356>.

- Smith, Bruce, "Scene," in: Henry S. Turner (ed.), *Early modern theatricality*, Oxford 2013, 93–112.
- Toutanova, Kristina/Klein, Dan/Manning, Christopher et al., "Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network," in: *Proceedings of Hlt-Naacl* (2003), 252–259.
- Trilcke, Peer/Fischer, Frank/Kampkaspar, Dario, "Digital Network Analysis of Dramatic Texts," in: *DH2015 Conference Abstracts*, Sydney 2015, <https://dina.github.io/presentations/2015-sydney/sydney.html/>.
- Trilcke, Peer/Fischer, Frank/Göbel, Mathias et al., "Netzwerkdyamik und Plotanalyse. Zur Visualisierung und Berechnung der 'progressiven Strukturierung' literarischer Texte," in: *Book of Abstracts of Dhd 2017*, Bern 2017, 175–180
- Underwood, Ted/Goldstone, Andrew, "What can topic models of PMLA teach us about the history of literary scholarship?" in: *Journal of Digital Humanities* 2/1 (2014), <http://journalofdigitalhumanities.org/2-1/what-can-topic-models-of-pmla-teach-us-by-ted-underwood-and-andrew-goldstone/>.
- Vogel, Juliane, "Aus dem Takt: Aufttritsstrukturen in Schillers *Don Carlos*," in: *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* 86/4 (2012), 532–546.
- Weinrich, Harald, "Hoch und Niedrig in der Literatur," in: Horst Rüdiger (ed.), *Literatur und Dichtung*, Stuttgart u.a. 1973, 160–170.
- Willand, Marcus, "Hermeneutische Interpretation und digitale Analyse. Eine Verhältnisbestimmung," in: Luisa Banki (ed.), *Lektüren. Positionen Zeitgenössischer Philologie*, Trier 2017, 77–100.
- Willand, Marcus/Reiter, Nils, "Geschlecht und Gattung. Digitale Analysen von Kleists *Familie Schroffenstein*," in: *Kleist-Jahrbuch* (2017), 177–195.

Digital Resources

Apache UIMA, <http://uima.apache.org>.

GitHub, QuaDrama, <https://quadrama.github.io>.

TextGrid Repository, <https://textgridrep.org>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Operationalising the Change. Dispersion of Polish literary life (1989–2002)

Maciej Maryl

1 Introduction

Let us picture history of literature as an unknown city one tries to explore. An individual may contemplate some details and visit places recommended by others. But one needs to climb higher to perceive regularities and connections between landmarks, to see hidden patterns of the city life. While climbing higher and higher we lose many details and subjective impression of particular buildings, yet, we get a fuller picture, and discover new relationships, invisible from the ground level. But you can never climb high enough to get a full picture. Or can you? This is precisely what this project aims to explore—the possibilities and limitations of a bird’s eye view on events, works, and processes in an important period of Polish literary history and comparing it with intersubjective claims by various critics and scholars.

Why the intersubjective view is not enough? All syntheses in literary history, as well as in other branches of humanistic inquiry are by necessity based on certain approximations due to obvious cognitive limitations of mortal, selective and forgetful human beings. How about expanding the capacities of individual researchers by providing them with the opportunity to read all the books at once, to learn all the facts (including the minor and seemingly unimportant ones) and combine everything into a meaningful theory? If we consider digital research tools not only as more effective typewriters, but—in McLuhanian terms—as extensions of our senses and capacities, we could treat such a ‘view from above’ as a great opportunity to generate new insights that could support our interpretative practices. Moreover, apart from cognitive limitations we may go beyond the set

M. Maryl (✉)

Digital Humanities Centre, Institute of Literary Research of the Polish Academy of Sciences,
Warsaw, Poland

E-Mail: maciej.maryl@ibl.waw.pl

of works considered canonical and explore also those objects which remained unnoticed by contemporaries or next generations of critics. According to laws of ‘historical selection’, proposed by Robert Escarpit, 80% of literary production becomes forgotten within a year and 99% in twenty years¹. By looking ‘from above’ we may now account for those forgotten works and try to present a more accurate picture of literary processes during a given period.

Digital revolution in the humanities brought about not only the increased storage capacity and accessibility of literary resources but more importantly the possibility to reuse them for the large-scale quantitative research. In doing so we may take the advantage of existing resources that have been diligently collected and annotated in various institutions throughout the years with the intention to facilitate the research process: bibliographies, catalogues, chronicles, calendars, fact sheets, reports, etc. There is a growing body of new, computational methods and approaches to the study of literature, which have clearly marked the distinction between traditional, close-reading and detail-oriented approaches, on the one hand, and new, synthetic studies applying statistical methods to larger corpora, on the other. In order to highlight the qualitative and empirical angle of this approach, it is dubbed “distant reading”², “algorithmic criticism”³, or “macroanalysis”⁴. The main premise behind this research corresponds with the metaphor I used in the introduction: when looking ‘from above’ we may lose some granularity concerning individual works, but instead we are able to survey larger-scale processes, sequences and cycles. Among these digital bird’s eye methods, we may distinguish between the scholarship analysing the features of actual texts and metadata-driven research.

The former approach concentrates on quantifying stylistic features of texts. For instance, stylometric research focuses on the most frequent words in a given corpus of texts.⁵ Such research may entail authorship attribution⁶, genre analysis⁷ or large-scale analyses of national literary output⁸. Stylometric analysis is also

¹ Escarpit, Robert. 1966. *The Book Revolution*. Toronto: Harrap.

² Franco Moretti, *Distant Reading*, London 2013a.

³ Stephen Ramsay, *Reading Machines. Towards an Algorithmic Criticism*, Urbana 2011.

⁴ Matthew L. Jockers, *Macroanalysis. Digital Methods & Literary History*, Chicago 2013.

⁵ Eder, Maciej/Kestemont, Mike/Rybicki, Jan, “Stylometry with R: a suite of tools,” in: *Digital Humanities 2013: Conference Abstracts*, University of Nebraska-Lincoln, NE (2013), 487–489.

⁶ E.g. Maciej Eder, “Style-markers in authorship attribution: a cross-language study of authorial fingerprint,” in: *Studies in Polish Linguistics* 6 (2011), 99–114.

⁷ E.g. Christof Schöch/Steffen Pielström, “Für eine computergestützte literarische Gattungsstilistik,” in: *1. Jahrestagung der Digital Humanities im deutschsprachigen Raum (DHd)* (2014), http://dig-hum.de/sites/dig-hum.de/files/Schoch-Pielstrom_2014_Gattungsstilistik.pdf (retrieved on April 20, 2018).

⁸ E.g. for German literature: Fotis Jannidis/Gerhard Lauer, “Burrows’s Delta and Its Use in German Literary History,” in: Matt Erlin/Lynne Tatlock (eds.), *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*, New York 2014, 29–54; for Polish novels: Jan Rybicki, “Pierwszy rzut oka na stylometryczną mapę literatury polskiej,” in: *Teksty Drugie* 2 (2014), 106–128.

combined with functional or grammatical linguistic categories for the study of prose⁹ or other literary genres¹⁰.

The second strand of research employs metadata, i.e. data about texts like title, genre, author, publication date, etc. Although such research often draws from the book history, current scholarship proves its validity for literary studies. Franco Moretti visualized and interpreted genre life-cycle on the example of metadata of the 19th-century British novels.¹¹ Matthew Jockers used bibliographical data to explore history, topics and geographical distribution of Irish American fiction.¹² Katherine Bode used various bibliographical datasets to “recalibrate” the history of Australian literature and challenge some established arguments including views on relationships between local and foreign publishing markets or gender distribution among authors and their coverage in press and scholarship.¹³ Through visualisation of data on poet-journal affiliation Long and So pictured the differences between literary networks in the U.S., Japan and China.¹⁴ It should also be noted that the division between linguistic and metadata approach is purely conceptual and they could be incorporated into a single research design, as in the case of Moretti’s “Reflections on 7000 titles.”¹⁵

This paper explores a possibility of quantitative research into transitions in literary history on the example of Polish literary life 1989–2002. This period is by no means under-researched. On the contrary, the abundance of scholarly and critical writing makes it even more interesting from the data-driven research perspective, since it allows for validating critical claims against existing data.

This exploration will be carried out on the basis of the data from *Polish Literary Bibliography (PBL)*¹⁶, a comprehensive database of Polish literary and cultural life. It indexes not only literary books but all instances of literary life and reception: reviews, journal articles, newspaper mentions, dramas, adaptations of literary works (stage, film, radio and television) and literary prizes. So, not only do we have access to rich data about literature but also to relationships between the objects (e.g. A is a review of B; C is an adaptation of D). The content is also annotated according its subject. This resource has been developed since 1954 and already covers Polish literary culture from 1944 to 2002. Unfortunately,

⁹Sarah Allison/Ryan Heuser/Matthew L. Jockers et al., “Quantitative Formalism: An Experiment,” in: *Pamphlets of the Stanford Literary Lab* 1 (2011), <https://litlab.stanford.edu/LiteraryLabPamphlet1.pdf> (retrieved on April 20, 2018).

¹⁰David L. Hoover/Jonathan Culpeper/Kieran O’Halloran, *Digital Literary Studies: Corpus Approaches to Poetry, Prose, and Drama*, New York 2014.

¹¹Franco Moretti, *Graphs, Maps, Trees: Abstract Models for a Literary History*, London 2005.

¹²Jockers (Ann. 3), 35–62.

¹³Katherine Bode, *Reading by numbers. Recalibrating the literary field*, London 2014.

¹⁴Hoyt Long/Richard So, “Network Science and Literary History,” in: *Leonardo* 46/3 (2013).

¹⁵Moretti (Ann. 1), chapter “Style, Inc.: Reflections on Seven Thousand Titles (British Novels, 1740–1850),” 134–158.

¹⁶<https://pbl.ibl.waw.pl/> (retrieved on April 20, 2018).

most of the data are still available only in print, with over 1 million digital records covering only years 1988 to 2003. The recent development project¹⁷ was aimed at converting the Oracle database of *PBL* into a new, more accessible format with some tools for researchers. The project also entailed a prototype for retrodigitisation of older volumes but, given the vast scope of the task, it is hard to foresee, when the full database would be available.

Hence this paper consists of two interconnected goals: (1) to test selected qualitative hypotheses concerning Polish literature of the transition period (1989–2002); (2) to propose methodology for the data-driven study of literary history with the use of bibliographical datasets. Both goals will be supplemented with a reflection on how documentation methodologies and practices affect the quantitative analysis.

2 Literary transition as a challenge to literary theory

As Anna Nasiłowska observes the periodization of the 20th-century Polish literature is almost never based solely on artistic criteria and was heavily influenced by external—political and economic—events.¹⁸ This is one of the reasons why she chooses 1976 as the starting point of her book on contemporary Polish literary history—the year of massive workers’ strikes which were violently broken by the authorities, what served as decisive impulse for the establishment of first organized oppositional bodies and, in turn, underground publishing. Nasiłowska uses the term “divided literature” to describe Polish literature between 1976 and 1989, what captures interesting dualism of literary life after the birth of *drugi obieg*, i.e. the *second circulation*, independent from the authorities and censorship. On the one hand there was official circulation, comprising of state-controlled publishing and dissemination institutions, and on the other a variety of *samizdat* books and journals, illegally printed and distributed by various dissident bodies. 1976 marks the beginning of organized underground printing in Poland, but Polish literature had been already published by émigré circles since the end of World War II, mostly in France and the UK but also in smaller centres in Germany, USA, Canada, and Sweden. The publications of those presses were targeted at local Polish communities, yet they were also smuggled to Poland and, after the establishment of illegal presses, reprinted (and *vice versa*—Polish *samizdat* was often republished abroad). And there was also the *third circulation* (*trzeci obieg*), i.e. counterculture publications, which were also published illegally but remained critical of political affinities and goals of the underground presses.

¹⁷Maciej Maryl/Piotr Wciślik, “Remediations of Polish Literary Bibliography: Towards a Lossless and Sustainable Retro-Conversion Model for Bibliographical Data,” in: *Digital humanities 2016: Conference Abstracts*, Kraków 2016, 621–623, <http://dh2016.adho.org/abstracts/293> (retrieved on April 20, 2018).

¹⁸Anna Nasiłowska, *Literatura Okresu Przejściowego 1975–1996*, Warszawa 2006.

Defining literary transition poses a challenge to literary theory, because of the sheer fuzziness of the concept. First of all, it is difficult to set up a clear date of the political breakthrough—is it June 4th 1989, the day of the first partly free parliamentary election, or August 24th, when the first non-communist government was appointed, or, what is even more pertinent from the cultural point of view, April 11th 1990, the day when censorship was officially abolished? As Przemysław Czapliński notes, “Transition, whether we want it or not, lasts. It is not a rupture but a movement past a critical point [*przesilenie*], not an event but a process.”¹⁹ For this reason literary critics discuss their observations of literary change in a broader timeframe, which makes those elusive processes visible: Nasiłowska focuses on twenty years (1976–1996), she calls *transitory period* (*okres przejściowy*);²⁰ Czapliński writes about the *traces of the end* in the literature of the 1980s and *traces of the beginning* in the 1990s;²¹ Janusz Sławiński, writes about *disappearance of the centre* (*zanik centrali*) in 1994, retrospectively naming the processes which gradually took place in the decade 1983–1993.²² There always has to be a point of reference. It should be noted here, although I am addressing this problem a bit later in this section, that a demarcation of the transition period ending point poses similar challenges.

Secondly, in thinking about the literary transition of 1989 there is an enthymematic assumption that political events bring about a cultural change. Many literary critics, as well as the audience anticipated a transformed, brand new literature.²³ Those expectations are neatly captured as *appetite for change* in the title of Jerzy Jarzębski book (as well as in the book itself). Czapliński and Śliwiński note that the need for literary breakthrough was intensified by ongoing debates about transformations of the society and they conclude that “the question, whether a political change [*przełom*] triggers a literary one, is and is not unreasonable at the same time. It is unreasonable, because—luckily—literature is not an addition to political life. It is not, because literature is a part of reality, which underwent an enormous change.”²⁴ So, perhaps, political change should be perceived as a sort of catalyst of cultural transition, because “it was not only a beginning of serious and extensive transformations of the infrastructure of the literary life, but it also ... made the novelty legitimate, because it built up the expectation of novelty.”²⁵

¹⁹Przemysław Czapliński, *Ruchome Marginesy: Szkice O Literaturze Lat 90*, Kraków 2002, 6. All translations from the Polish sources are by the author.

²⁰Nasiłowska (Ann. 17).

²¹Przemysław Czapliński, *Ślady Przełomu*, Kraków 1997.

²²Janusz Sławiński “Zanik centrali,” in: *Kresy* 18 (1994), 14–16.

²³Cf. Przemysław Czapliński/Piotr Śliwiński, *Literatura Polska 1976–1998: Przewodnik po prozie i poezji*, Kraków 1999, 209–213; Czapliński (Ann. 18), 11–24.

²⁴Czapliński/Śliwiński (Ann. 19), 213.

²⁵Ibid., 113.

In his book exclusively dedicated to the transition, Czapliński notes that literary breakthrough takes place in three spheres of poetics, ideas and institutions.²⁶ It affects poetics, when certain genres and conventions lose their function of carriers of important content and new, different forms replace the accepted ones.²⁷ The field of ideas undergoes change thanks to new modes of using literature.²⁸ Finally, the institutions are transformed when the means of literary production and distribution are changed, together with relationships between the partners of literary communications: writers, patrons, publishers, educators.²⁹ Given the vagueness of the plane of ideas (and, to some extent, the poetics), the change is mostly visible in the institutional field, Czapliński concludes.³⁰ In this paper I try to map the evolution of literary institutions from the centralised model to the commercialised literary market of the late capitalism.

‘Mapping’ is not an accidental word here, since critics often used spatial metaphors to describe the processes of transition. The concept of *disappearing centre* was coined by Sławiński, an eminent Polish structuralist, in a tiny (3-pages long) and highly influential essay.³¹ Although he actually focused solely on poetry, his remarks were applied, with much success, to literary life in the broadest sense. Sławiński’s point of departure is a vision of literature in which individual works are parts of the *greater Whole* (*większa Całość*), in which *authoritative* (*miarodajni*) poets (i.e. those “read by everyone, whether they like them or not”) serve as orientation points for the assessment of other works.³² In other words, they constitute the canon, in relation to which the remainder of literature could be positioned. Sławiński observes the collapse of this system and concludes:

This Whole, which had supplied common frames of literary communication and had organised it around a certain centre, has been replaced by a polycentric system: multiplicity of—social and spiritual—micro worlds, in which poets and readers find each other. In each of them there are local tastes, local evaluation measures, local circulations, local relationships, authorities and hierarchies, local lucidities and unlucidities. One writes mostly for fellows ... and expects new poems—from fellows.³³

This spatial metaphor is eventually elaborated by Bałajewski, who claims that literature is in the *state of dispersion* (*stan rozproszenia*) by demonstrating how top-down mechanisms of literary prestige are being substituted by bottom-up canon formation: locally acclaimed writers, who write about their local

²⁶ Ibid., 5.

²⁷ Ibid.

²⁸ Ibid.

²⁹ Ibid.

³⁰ Ibid., 5.

³¹ Janusz Sławiński, “Zanik centrali,” in: *Kresy* 18 (1994), 14–16.

³² Ibid., 15.

³³ Ibid., 16.

experience gain national recognition.³⁴ ‘Locality’ in his understanding goes beyond geography, covering like-minded small initiatives in different parts of the country.³⁵ What is important in this diagnosis is the feeling of temporariness: “Situation is still quite fluid. New initiatives appear all the time, transforming a—seemingly—ready cultural-institutional-literary system. We are in the state of dispersion.”³⁶ This flexible relationship between the centre and peripheries will be later elaborated by Czapliński, who describes this process through a metaphor of *moveable margins* (*ruchome marginesy*).³⁷ It is a vision of literary culture in which different phenomena coexist and either move to the centre or are pushed to the margins, depending on their importance for the greater community.

Those transformations are, if not welcome, at least accepted by the critics who—perhaps sharing the *appetite for change*—embrace this new configuration of cultural production, acknowledging its potential. Yet, quite quickly this optimism yields to disappointment, mostly due to the fact that decentralisation processes meet the market economy. Underfunded local initiatives find themselves on the margins, whereas popular and commercial literature occupies the centre. “Thanks to the domination of the market ... the centre has become established and the margins fixed.”³⁸

Somewhere in the second half of the 1990s the trend is reversed, marking the beginning of a slow normalisation, understood as a process in which literature is marginalised by mass media, what in turn is treated as a natural order of things by the public.³⁹ Critics start challenging the dominant discourse which establishes new, commercially-generated hierarchies. They call for a literature which would challenge the readers and face them with otherness.⁴⁰ Those processes coincide with the *return of the old masters*,⁴¹ i.e. revaluation of the great writers who are again featured in mainstream and receive literary prizes. In the end, critics claim, we are witnessing the *return of the centre* (*powrót centrali*),⁴² but the centre has changed. It is not the official, state-controlled system, but the dominant media discourse⁴³ and free market processes. Hence, the state of dispersion is considered a temporary stage between two centralized systems.⁴⁴ As Czapliński claims, the

³⁴ Arkadiusz Bagłajewski, “Stan rozproszenia,” in: *Odra 1* (1996), 51–55.

³⁵ *Ibid.*, 19.

³⁶ *Ibid.*, 17.

³⁷ Czapliński (Ann. 18).

³⁸ *Ibid.*, 7.

³⁹ Kinga Dunin, “Normalka,” in: Przemysław Czapliński/Piotr Śliwiński (eds.), *Normalność i konflikty: rozważania o literaturze i życiu literackim w nowych czasach*, Poznań 2006.

⁴⁰ Czapliński (Ann. 18), 21–36.

⁴¹ Arkadiusz Bagłajewski, *Mapy Dwudziestolecia 1989–2009: Linie Ciągłości*, Lublin 2012.

⁴² Przemysław Czapliński, *Powrót centrali: Literatura w nowej rzeczywistości*, Kraków 2007.

⁴³ Dunin (Ann. 35).

⁴⁴ Bagłajewski (Ann. 37).

only difference between those two being that the former was politically-driven and the latter is established on commercial principles.⁴⁵

Let us conclude with presenting the claims of the critics as set of hypotheses which we could try to verify on the basis of available data. Firstly, the institutions of literary life undergo dispersion in the early 1990s, followed by its subsequent re-centralisation according to the market principles. Secondly, literary reception follows the same pattern. Thirdly, literary reception in the ‘dispersed’ phase is debutocentric (i.e. focused on debutants), which changes in the re-centralisation phase with more attention being given to the ‘old masters,’ that is, eminent writers who earned their position before 1989. Finally, the processes of dispersion and re-centralisation also have a geographical dimension, which should be visible as the rise, and eventual decline, of smaller publishing centres.

3 Operationalisation

The first task is to translate elusive metaphors into variables or indicators. As Franco Moretti puts it in his pamphlet on character-space: “Operationalizing means building a bridge from concepts to measurement, and then to the world. In our case: from the concepts of literary theory, through some form of quantification, to literary texts.”⁴⁶ In our case: from the concepts of literary history, through some form of quantification, to bibliographical data.

Apparently this task looks like an equation with two unknowns: on one hand there are metaphoric claims, on the other bibliographical material prepared by humans, full of possible errors and biases. However, in a spirit most recently advocated by Kathrine Bode, we should use the “bibliographical consciousness”⁴⁷ and recognize the shortcomings of the data, what in turn will allow for a “data-rich literary history”, accounting for the phenomena which are graspable with this dataset (and such an account is given later in this section). But how should we approach the other unknown, namely claims made by literary critics of the transition period? Firstly, there is a question whether those generalisations are of a descriptive or rather prescriptive nature, as critics are *nolens volens* actors of the literary life they investigate.

Moreover, one should not overlook the rhetoric glare of many metaphors coined by critics. Metaphors capture the ideas in an attractive manner but also sparkle the debate on their validity. One of the critics, Przemysław Czapliński, provides a perfect material for operationalisation, because he has coined and

⁴⁵ Czapliński (Ann. 38).

⁴⁶ Franco Moretti, “‘Operationalizing’: or, the function of measurement in modern literary theory,” in: *Pamphlets of the Stanford Literary Lab* 6 (2013b), 3, <https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf> (retrieved on April 20, 2018).

⁴⁷ Katherine Bode, “The Equivalence of ‘Close’ and ‘Distant’ Reading. Or, toward a New Object for Data-Rich Literary History,” in: *Modern Language Quarterly* 78/1 (2017), 77–106, here: 94.

defined many concepts concerning the period in question. Yet, his critical strategy was also criticised for this very reason: “He identifies, reads, recommends, diagnoses and describes trends, styles, poetics and circumstances but he does not show the path for literature because, despite his outstanding intellectual resourcefulness, he never gets attached to his ideas, proclaiming the new ones all the time. Fast enough he forgets about the previous ones.”⁴⁸

Unlike critics, who form their judgments on the basis of vast cultural material (although accessed unevenly and unsystematically), empirical approaches are faced with a scarcity and fragmentation of the data at hand. Unlike scholars in life sciences, who collect data relevant for their hypotheses, we need to tune our research scope to available material. Hence the operationalisation becomes a tricky activity which entails “building a bridge from concepts to measurement” with only that many bricks we can gather.

The most important thing to be noted about the *PBL* data is that it itself is a construct guided by certain methodological principles that has evolved overtime. Luckily the period I am working with would not witness major methodological changes, however there are important reservations to be made. First of all, although comprehensive and detailed, the *PBL* database does not contain every piece of literature published in Poland. For instance, it does not account for some regional literary journals. It may also favour certain eminent writers by collecting all their output, even of clearly non-literary nature. Secondly, literary documentation, unlike literary criticism, has to work with a sharp definition of literature. Hermeneuts can dedicate entire volumes to discussions of literariness, whereas literary documentation has to operate within well-defined borders. The consequences are grave—if something is considered literary it belongs in the database.⁴⁹ Thirdly, only certain data fields (creator, reviewer, publisher, journal, particular work) have controlled vocabularies, and some entities have duplicate entries. I tried to minimise this influence in the data-cleaning process, however there may be human mistakes I overlooked. Moreover, an open ‘comment’ field for each data entry contains lots of unstructured information (e.g. if a critical work discusses two authors, the less featured one will be mentioned only in annotations, thus not available for quantitative analyses). Ongoing modernisation of the *PBL* will minimise those problems in the future.

With all those reservations in mind, we may treat the *PBL* database as a broad, representative sample of Polish literary life, which is carefully annotated and ready for the analysis. It is safe to assume that although the absolute results obtained on this data may not be comparable with other resources (e.g. actual number of books published or creators active in the given year), yet thanks to

⁴⁸Stanisław Gawliński, “Wstęp,” in: Stanisław Gawliński/Dorota Siwor (eds.), *Dwie dekady nowej (?) literatury 1989–2009*, Kraków 2011, 79.

⁴⁹Cf. Tomasz Umerle, “Rethinking the Potential of Documentation of Culture as a Data Gathering Practice,” in: *Proceedings from the Document Academy 4* (2017), <https://doi.org/10.35492/docam/4/2/15>.

consistent methodology it allows for relative comparisons overtime, i.e. looking at certain trends in literary production.

The initial dataset was downloaded from the *PBL* database. The database is relational, so the materials requested consisted of all books and articles published, including secondary literature. Although *PBL* covers Polish literature from 1944, the online database contains only the data for 1988–2003 with some earlier works, if they were referenced in the modern material. At the time of the study year 2003 was still not finished and year 1988 was not complete, as it was used as a test year for the setup of the first online database in 2000.⁵⁰ Given those constraints I limited the material to years 1989–2002. Furthermore, as the collection methodology was changed after the year 1989 was entered into the database, most notably through excluding some regional journals, I decided to exclude from the sample all journals which appear only in 1989.

Limiting the transformation period to 1989–2002 is an arbitrary move, dictated simply by the availability of the research data for comparative analysis. However, basing on the critical assumptions discussed above, one should be able to trace the processes of dispersion and subsequent re-centralisation in this period. Ideal scenario would require a dataset spanning approximately five years more in both directions. Hopefully, in the future data would allow for re-examination of the findings presented below.

There are three major datasets I use in this study:

- a) **Books.** Literary books either published in Poland (incl. translations of foreign literatures), or by Polish authors abroad. Each entry includes information about the author, publisher, place, and year of publication. In some cases book genre was available or had been inferred from metadata or the title (e.g. play, children literature, music, nonfiction, short stories, novel, reportage, art, poem, memoir).
- b) **Articles.** Literary articles published in Polish literary journals, including poems, fiction, letters, essays, nonfiction and adaptations. Metadata cover information about the author, journal, year and the place of publication (in some cases).
- c) **Reception.** Various kinds of secondary literature, i.e. review, interview, summary, letter, polemic, etc. Each piece of reception is connected to one literary author (creator). Some reviewers were also authors but in this study I omitted this relationship. Certain reception pieces noted in the bibliography are about non-literary books and authors (e.g. review of a literary study).

For all datasets I prepared tables allowing for further computation. This entailed data cleaning (esp. removing typos and mistakes) and enriching with additional information (e.g. author's gender, latitude and longitude of cities). Certain

⁵⁰Certain efforts are being made towards digitising printed volumes and hence expanding the range of the online bibliography.

variables, have low reliability due to changing descriptive conventions and human errors, e.g. number of pages, or journal issue numbers.⁵¹

Altogether the dataset consists of 33,142 literary books and 138,925 literary articles by 21,865 authors. These were published by 8233 publishers and 958 journals. There are 93,874 reception texts, including pieces about 9286 authors and 22,335 books.

4 Studies of dispersion

The movement from dispersion towards centralisation and back, as described by critics, could be captured on many levels. The most obvious one, is simple bibliometric statistics, which could reveal the structure and evolution of literary life over the transformation period. I commence with some statistics of the literary production of that period, then combine those findings with the reception data, and subsequently, with comparison of reception of debutants and ‘old masters’. Next I analyse the dispersion on the example of geography and finally use network analysis for centrality measures.

4.1 Literary production

Let us start with sheer numbers concerning the literary production of that period. Dispersion, operationalised on the available dataset, would entail an increase in number of smaller publishers, journals or publishing centres disseminating literary content over time. Recentralisation would mean the reverse, i.e. smaller number of big players overtaking literary communication. Let us see how this look through some of the available measures.

To put the discussion in context we should start with a general statement that the period in question in Polish literature could be characterised in terms of growth. Figure 1 combines yearly numbers of publications (books and journal pieces), active literary journals (in this example only those which published literary works, not reception pieces), as well as count of publishers and authors (i.e. those who published at least one book or a journal piece). Trend lines show a steady growth of two first groups and quite a stable situation of the last one. This tendency is not so obvious if we take into consideration economic factors (diminishing of state-subsidized publishing) as well as the competition of other media in 1990s.

In order to add some granularity to those frequencies I divided publishers and journals into big and small. If critical assumptions were correct we should see lots of small, dispersed presses in the early 1990s and subsequent market

⁵¹ I am very thankful to Cezary Rosiński from the *PBL* team, who extracted the data and prepared a list of debutants, referred to in the next section.

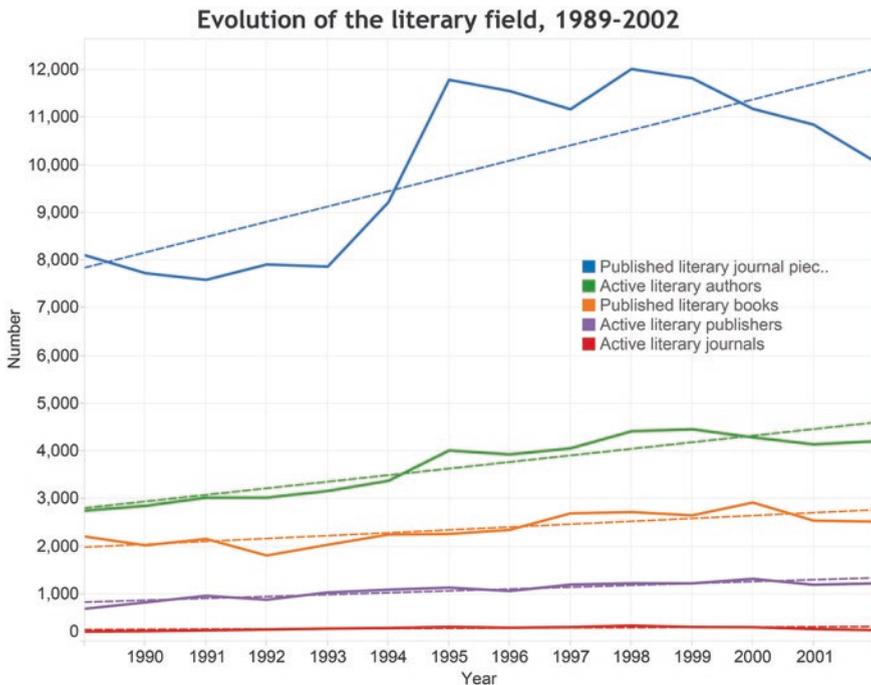


Fig. 1 Evolution of the literary field

re-concentration towards the turn of the decades. In order to differentiate the publishers and journals I took their annual publication numbers and applied the 90th percentile as the threshold for qualifying a journal or publisher as big or small. In other words, big players were 10% of the most prolific journals and publishers each year. It should be noted that the quality of being “big” or “small” is always relative to a single year and the publishing numbers of other entities; a publisher or journal with the same number of publications over time may be qualified as a big one in one year, and as a small one in the other, depending on the outcomes of other publishers.

Area graphs in Fig. 2 show interesting discrepancy between the actual number of journals and publishers and their average output, i.e. average number of publications by active entities in all four categories. The production seems to be dominated by large players, which is especially true for big journals after 1995 and big publishers since around 1998. Interestingly, this growth of big publishing coincides with visible drop of the number of active players in the early 2000s.

It should be noted that the overrepresentation of small literary players may be partially caused by the data collection methodology, which would not register non-literary books. For instance, if a big publisher in the field of history releases one volume pertinent for literary studies (e.g. collection of letters by a literary figure), only this book will be counted and the publisher would appear in *PBL* as a small one.

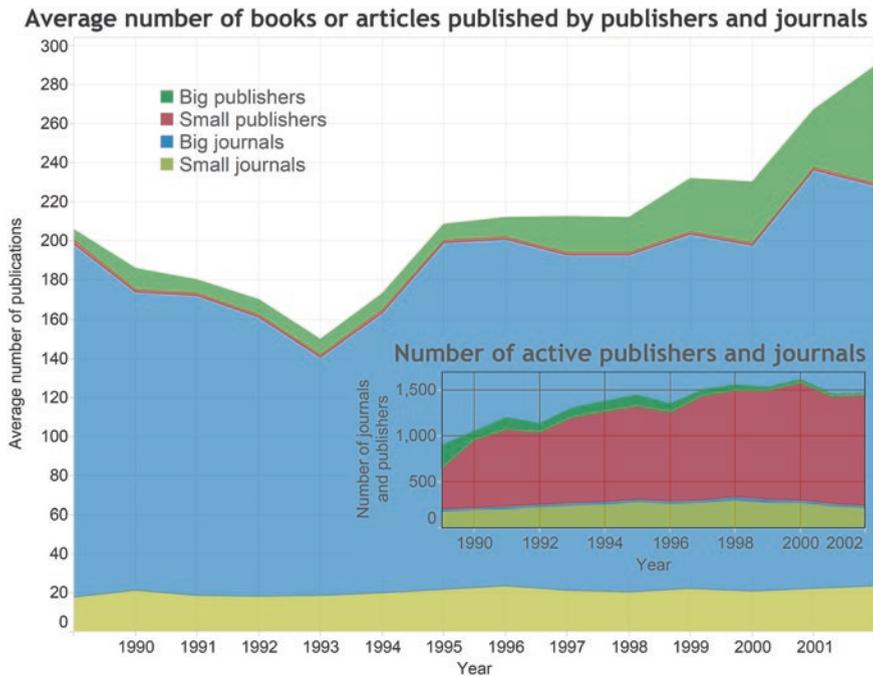


Fig. 2 Average number of books or articles published by publishers and journals

4.2 Critical reception

Sheer production statistics could be misleading, if we fail to take into consideration the reception practices. If we want to analyse the tendency for centralising the literary taste, we need to compare the number of books with its reception, namely how many times certain book is reviewed, discussed, or mentioned. Data in Fig. 2 seem to at least partially corroborate the claims of literary critics about eventual recentralization of literary life by a few bigger players. This trend is even more visible if we take into consideration the critical attention. Figure 3 plots the number of reviews of books by big and small publishers.

The critical attention devoted to books mirrors the processes described by the critics. On the centralised market of the 1989 Poland there was a clear domination of books by big publishers, which attracted most of the critical attention. Yet, they are swiftly outpaced by small publishers after 1991. This trend quickly reversed around 1995 and within two years big publishers regained the dominance in the struggle for critical attention. This trend is perhaps even more strikingly visible when compared with the actual publishing output.

In Fig. 4 the books by big (blue) and small (red) publishers are contrasted with the ratio of critical attention (green line represents the number of book reviews

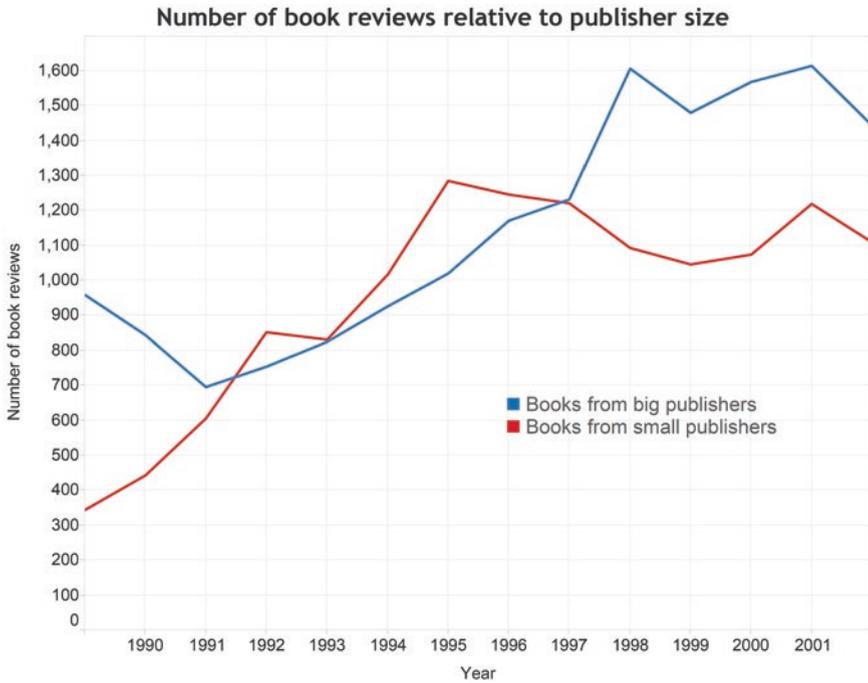


Fig. 3 Number of reviews dedicated to books by big and small publishers

of publications from small publishers divided by reviews of big publishers' output; score above 1 means higher interest in small publishers, whereas value below 1 indicates that critical attention shifts towards big players). It is worth noting that 1989 was the only year when big publishers published more than the small ones (55%). In the entire period small publishers had published on average 57% of all books (median = 58%), and in 2001 and 2002 the record numbers of 58% and 62% of all books published respectively. This pattern is compatible with the critical narrative about the recentralisation and the emergence of mainstream in the late 1990s.

In order to better understand those relationships we apply some measurements from economy.⁵² In this case a particularly useful is *Gini coefficient*, usually used to measure inequality in societies. We may consider literary life as a system in which attention is given to authors in three ways: by publishing their books, shorter pieces in journals, and by writing about their work or themselves. Measuring the distribution of this attention allows for judging the level of equality of writers according to those variables and for making further inferences about the

⁵²I am grateful to Mark Algee-Hewitt and Ryan Heuser for pointing me in that direction.

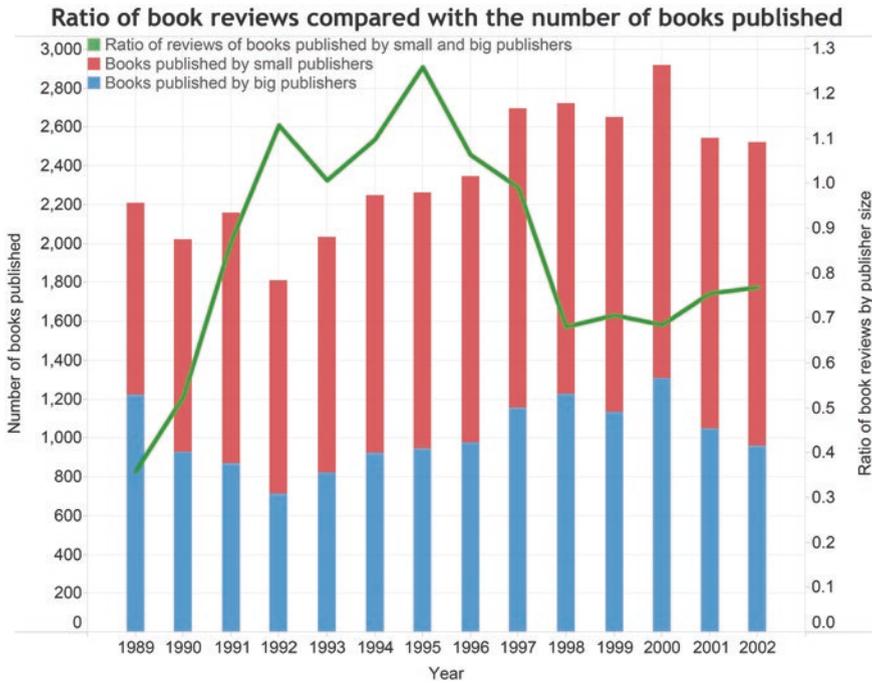


Fig. 4 Ratio of book reviews of publications from small and big publishers compared to their output

processes of literary life. Data for computation was prepared for three variables (books published, journal pieces published, personal reception pieces about the author) for every year between 1989–2002. Authors not present in a given year were omitted. *Gini coefficient* was computed in *ineq* package for R.⁵³ The results are presented in Fig. 5.

Gini coefficient takes a value between 0 and 1 and indicates the distance between actual distribution (as represented by Lorenz curve) and the perfectly equal one. Hence, the value 0 indicates perfect equality (all authors publish the same number of books), and 1 the opposite, absolute inequality (one author publishes all the books). First, it should be noted that it is hardly surprising that literary culture seems to be unequal, yet what strikes in Fig. 5 is the discrepancy between the distribution of the books published and other variables. How to interpret it? If we look at the World Bank statistics of economical distribution in various countries, we will notice that even most unequal societies rarely score higher than 0.5. Through this comparison we may conclude that distribution

⁵³Achim Zeileis, *ineq: Measuring Inequality, Concentration, and Poverty*, R package version 0.2–13 (2014), <https://CRAN.R-project.org/package=ineq> (retrieved on April 20, 2018).

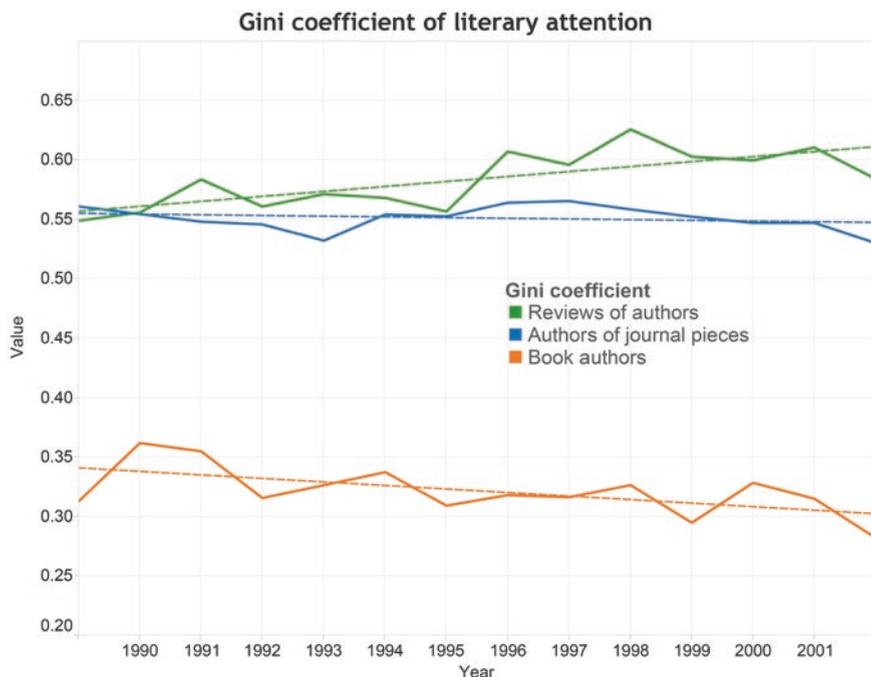


Fig. 5 *Gini coefficient of literary attention*

of critical attention and journal pieces is highly unequal, whereas the authors publishing books are represented more equally. Moreover, the publishing data show a slightly downward trend, whereas the inequality of critical attention is on the rise and seems to follow the pattern visible in previous figures (sharp rise in the second half of the 1990s and a drop in the last year). This is by far the only variable in this figure, which follows the claims of the critic and seems to point out to some re-centralisation processes in the field of secondary literature. Let us examine it more closely, through examination of another metaphor describing the transformation—the return of the ‘old masters’.

4.3 Debutants and masters

One of the critics claimed that in the first part of the 1990s we witnessed *debutocentrism* in Polish literature, i.e. concentration on debutants, which eventually shifted to the interest in the “old masters”, i.e. eminent writers, who earned their position before 1989. In order to assess this claim the notions of debutants and masters were operationalised. Debutants are those authors, who have not published anything before 1989, whereas the masters were selected qualitatively as those authors, who were highly popular before 1989 and featured

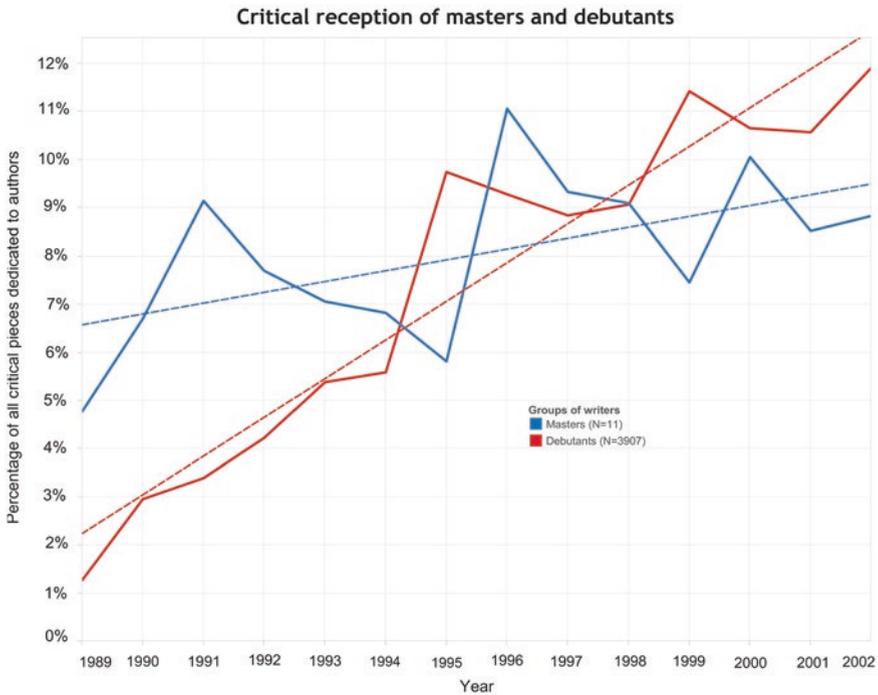


Fig. 6 Critical reception of masters and debutants

in criticism (e.g. Czesław Miłosz, Wisława Szymborska, Tadeusz Różewicz, Zbigniew Herbert, Stanisław Lem). The third, largest group of other writers (including historical figures) was left out. This selection led to establishing a two groups consisting of 11 masters and 3907 debutants. The reception of both groups is plotted in Fig. 6. The y axis represents the percentage of critical texts dedicated to authors from both groups, relative to all articles about authors published in a given year. Yet, despite similarities of scores, we must not forget that the groups compared are extremely different in size.

We may start with the observation that trends in both cases point to increasing interest. Furthermore, although the steep climb of writings dedicated to debutants in the early 1990s may be an indicator of *debutocentrism*, we hardly see a decline in the 2nd part of the decade. Yet, this could be a side effect of operationalisation, which defines debutants as all authors, who published for the first time in 1989 or later. Hence the number of debutants increases overtime, what may cause the gradual rise of critical attention. Yet, if we concentrated on the trends in the first half of the 1990s we can actually detect a decline in interest in acclaimed writers, who are overtaken by debutants by 1995. Suddenly, this trend almost reverses between 1996 and 1998 and afterwards.

A closer scrutiny of the data allows to link those irregular spikes on the masters' line with particular events such as awarding Wisława Szymborska

with the Nobel prize in 1996, deaths of important writers—Zbigniew Herbert in 1998 and Gustaw Herling Grudziński in 2000, as well as with the 90th birthday of another Polish Nobel-prize winner, Czesław Miłosz in 2001. In all cases the abundance of writing about those authors would dramatically increase the overall reception of masters. And such events are not only limited to the living authors. For instance, the 200th birthday of Adam Mickiewicz in 1998, an eminent poet of Polish romanticism, generated over one thousand articles about him. It seems that the Nobel prize for Szyborska may have been a turning point in (or at least an influence on) Polish literary history in that respect that it renewed the interest in established writers. Yet, it should be highlighted that this pattern of interest seems to be incidental and related solely to big events regarding particular writers. Szyborska is a good example here—in the Nobel year there were 306 critical pieces about her, almost twice as much as during the preceding seven years combined (127). A year later, there were 180 pieces about her and in 1999 she was featured in roughly the same number of articles as in the year before winning the prize (18). This interest seems to be quite short-lived and seasonal, yet it may explain inequality patterns in critical attention of the late 1990s, when individuals were subjects of disproportionately high attention.

4.4 Geography

Processes of dispersion have also geographical dimension. If the critics were right, we should witness the rise and subsequent decline of smaller publishing centres throughout the transformation period. In order to map this process I compiled the information about the locations of publishers and journals. I excluded publishing locations outside Poland in order to minimise the number of accidental locations. For each of the 607 Polish cities I computed the number of books and articles published there each year. In order to locate larger centres I used the same procedure as in the case of publishers and journals, setting a high publishing level above the 90th percentile of all the city publications for a given year. It is important to add that levels were computed separately for books and journal pieces, hence, for instance, in terms of book publishing a given city could be considered large centre, and be simultaneously classified as small in terms of the journal pieces published (and *vice versa*). The classification of a given city may also differ overtime.

Figure 7 contains two representations of these data. The graph above shows the number of works published in cities, where blue colour signifies the high level of literary production in the city, and red represents low publishing level. The numbers are given for both books (squares) and journal pieces (ovals). In terms of publishing we see a steady rise of small publishing centres, which reach a similar level as the large ones in 2002. In terms of journals, we see a steady growth in both types of cities. The graph below uses the same features but arranges them according to the actual number of cities falling under given category. Here again journals behave differently as we see a small rise and subsequent decline of the number of smaller locations which host literary journals. In terms of books, both levels seem to rise and there are more locations with both smaller and larger output.

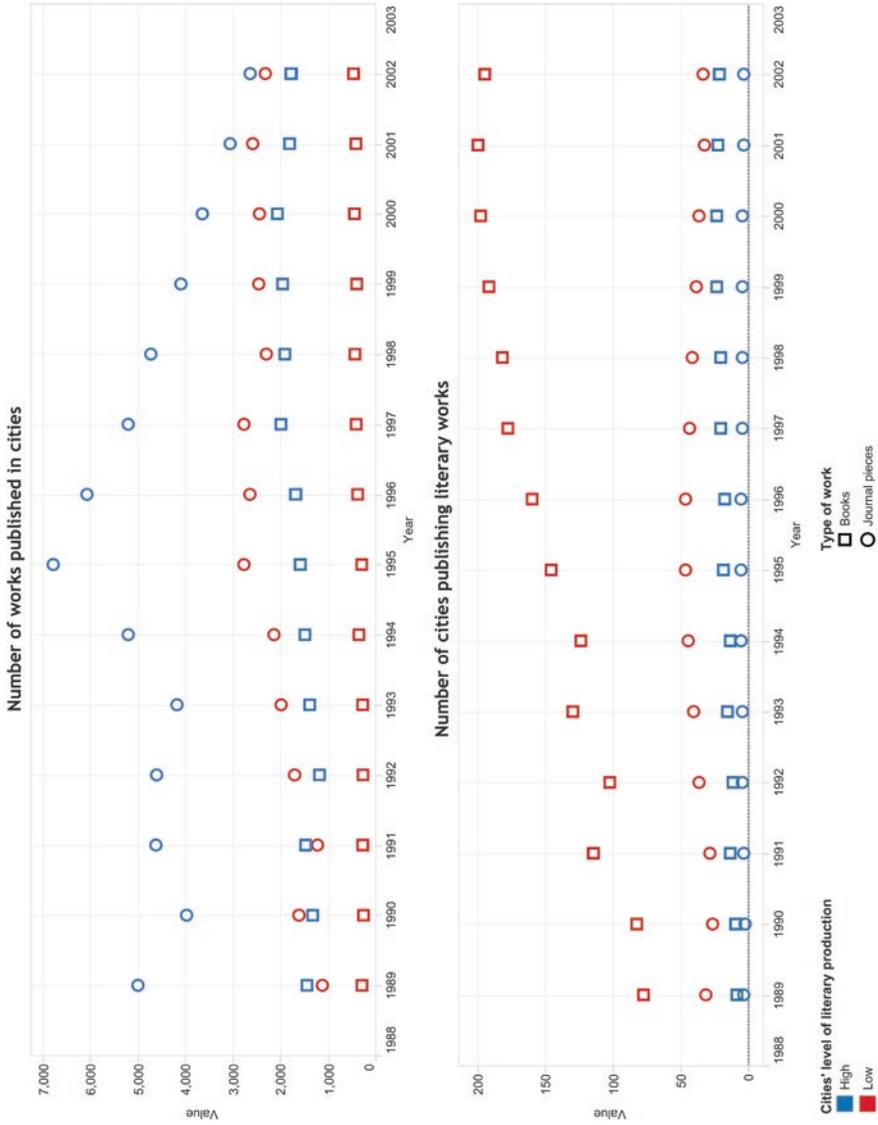


Fig. 7 Number of publications in Polish cities

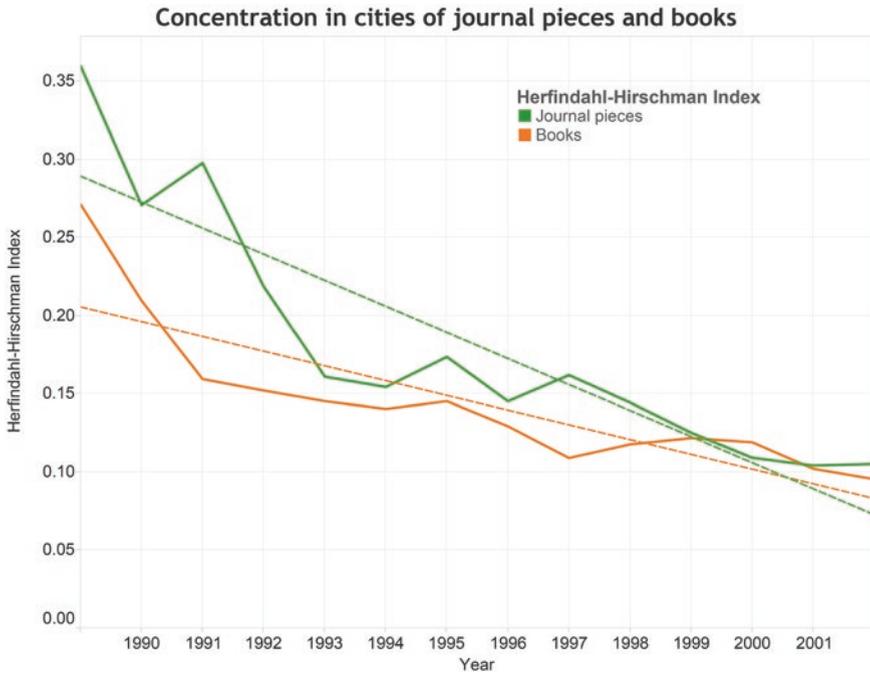


Fig. 8 Journal pieces and books concentration in cities

In order to shed more light on those numbers, let us once again turn to economy. *Herfindahl–Hirschman Index (HHI)* is usually used to describe market concentration, hence to answer the very question we are dealing with here. I computed HHI using *DescTools* package for R.⁵⁴ As in the case of *Gini coefficient*, described above, I generated a table containing the number of books and journal pieces published in a given city each year. Cities without publications in a given year were omitted. The results are presented in Fig. 8. *HHI* usually takes values from 0 to 1000, with the highest value representing monopoly. As it is a sum of squares of all entities' shares in the market we should not be surprised to find that distribution of literary production among few hundred cities is highly deconcentrated, because many cities have a very low publishing output (cf. Fig. 7). Hence, we should not concentrate on the absolute values, but rather on trends and change overtime.

HHI plot in Fig. 8 shows a clear trend towards deconcentration in terms of both number of books and journal pieces published in cities. It seems that almost each

⁵⁴Andri Signorell/Ken Aho/Andreas Alfons et al., *DescTools: Tools for Descriptive Statistics, R package version 0.99.24* (2018), <https://cran.r-project.org/package=DescTools> (retrieved on April 20, 2018).

year new actors emerge and take their small share of the market. Although we hardly see any trend towards re-centralisation, it should be noted that in both cases we can distinguish the radical deconcentration during the first 5 years, from the rather steady downward movement in the last five years. In between, i.e. around 1995, we even witness a small movement towards concentration.

4.5 Network analysis

So far I have been using various measurements of single variables to account for the processes in question. In the concluding study I borrow tools of network analysis to look at literary life as a network of relationships and interdependencies. On the basis of such relationships, like being published by the same publisher, covered in a journal, receiving the same prize, one can build a network of literary life and explore its shape, which could be centralised or dispersed.⁵⁵

The network of literary life that emerges from *PBL* data is complex because it features different types of actors and relationships between them. I decided to work with the author-centered network and linked them with other entities: publishers (published by), journals (published in, or reviewed in), reviewers (reviewed by). One can of course oppose to using reviews twice—to, first, mark the author's relationship with a reviewer and then with a journal—but it is providing an additional layer of connections (a reviewer usually publishes in other journals, and any journal usually publishes many reviewers). I did not link reviewers to journals (to keep the network author-centred) and I did not use cities (in order not to introduce another entity class). The network contains the total of 339,287 unweighted connections (including duplicate entries). For each year I prepared a list of relationships between authors on one side and other entities on another. Those tables were then transformed into lists of nodes and edges in *Gephi*.⁵⁶ Each node, i.e. relationship between two entities was assigned a weight, which indicates the number of times they were connected in a given year.

This network contains many different kinds of entities but I will simplify its interpretation and will treat it as a two-mode undirected network of writers and literary institutions. In order to deepen the understanding of this network I also use a one-mode 'projection' of this network, prepared with the use of bibliographic

⁵⁵Cf. Long/So (Ann. 13); Hoyt Long/Richard So, "Network analysis and the sociology of Modernism," in: *Boundary 2* 40/2 (2013), 147–82.

⁵⁶Mathieu Bastian/Sebastien Heymann/Mathieu Jacomy, "Gephi: an open source software for exploring and manipulating networks," in: *International AAAI Conference on Weblogs and Social Media* (2009), <https://gephi.org/publications/gephi-bastian-feb09.pdf> (retrieved on April 20, 2018).

coupling in Sci2.⁵⁷ A ‘projection’ is a translation of a two-mode network into relationships among the entities of one kind.⁵⁸ For instance, authors A and B, published novels with publisher X, and authors B and C printed their poems with publisher Y. In a two-mode network there are no connections between authors, but only between them and publishers (A–X, B–X, B–Y, C–Y) in a one-mode network relationships with publishers are translated into connections between authors on the basis of co-occurrence. So, B is connected to both A and C, but A and C are not connected directly because they did not publish together. Transforming network into a one-mode projection is often essential to proceed with network statistics as most measures are designed for this kind of networks.⁵⁹

The network concentration was measured with the average degree statistics. A node’s degree is the sum of its connections, while average degree is a mean of all degrees in the network. The concept of degrees differs in one- and two-mode networks. If a writer published only with two publishers, both of whom published books of 10 other authors, her two-mode degree would be 2 (2 publishers), while one-mode degree would amount to 20 (2×10 authors). Hence the latter is larger than the former. Figure 9 shows the average degree of one-mode and two-mode networks—the former was computed with *Gephi*, the latter, with *tmet* package for R.⁶⁰

For reasons described above, both networks operate on different scales but in the graph they were adjusted by the starting point, to allow for visual comparison. The one-mode network of relationships between writers (orange) seems to be perfectly fitting to the dispersion-recentralisation narrative, as after 1994 writers seem to increase their connections on average. Quite interestingly, the two-mode network (red), focused more on institutional relationships, tells a similar, yet not exactly same narrative. It seems as if it was more prone to the extremes (1993, 1996, 2001), as the general lower number of connections makes ‘stars’ more influential for the overall score. The sudden drop in the end corresponds with the general growth of book publishing, visible on earlier graphs. So, although big players may actually publish more books they remain unnoticed on the two-node network because of the wave of new publishers. The one-mode network rises on the contrary, probably because large publishers produce more books in general what allows for establishing new connections through co-occurrence.

⁵⁷ *Science of Science (Sci2) Tool. Indiana University and SciTech Strategies*, <https://sci2.cns.iu.edu>; cf. Shawn Graham, *Converting 2-mode networks to 1-mode networks* (2012), <https://electricarchaeology.ca/2012/02/08/converting-2-mode-networks-to-1-mode-networks/> (both retrieved on April 20, 2018).

⁵⁸ Cf. Tore Opsahl/Filip Agneessens/John Skvoretz, “Node centrality in weighted networks: Generalizing degree and shortest paths,” in: *Social Networks* 32/3 (2010), 245–251.

⁵⁹ Cf. Graham Shawn/Ian Milligan/Ian/Scott Weingart, *Exploring Big Historical Data: The Historian’s Macroscope*, London 2015, 195–234.

⁶⁰ Tore Opsahl, *Structure and Evolution of Weighted Networks*, University of London (Queen Mary College), London 2009, 104–122, unpublished thesis, <http://toreopsahl.com/publications/thesis/> (retrieved on April 20, 2018).

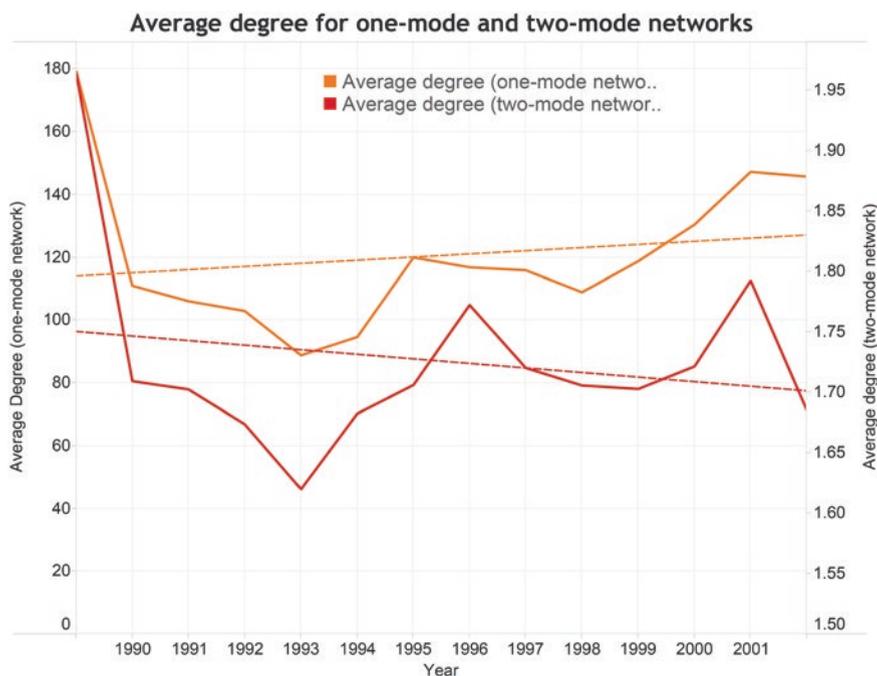


Fig. 9 Average degree of one- and two-mode networks

5 Conclusions

Probably the most obvious conclusion of this study is that there is no one clear way to operationalise the claims made by the critics. Once we start to disentangle the complex nature of literary life it appears that certain measures correspond with critical predictions and some do not. Interestingly the patterns which seem to corroborate the critical claims almost always contained the reception variable: patterns of critical attention given to big and small publishers, reception of debutants and old masters, *Gini coefficient* and the network analysis. The only time the decentralisation pattern was visible without the reception component was in the study of concentration of literary production with *HHI*.

Claims of the critics are based on various materials they have at hand: literary texts, critical discourse, scholarly articles. Some of those claims are based on empirical data.⁶¹ Their interpretation of macro processes is then an individual interpretation of the events. However, the sharper the diagnosis, the more question

⁶¹ E.g. Przemysław Czaplinski/Maciej Leciński/Eliza Szybowicz et al. (eds.), *Kalendarium Życia Literackiego 1976–2000: Wydarzenia, Dyskusje, Bilanse*, Kraków 2003, use number of journals having been shut down in the early 1990s, new literary prizes, number of debuts, etc.

it raises: what are the phenomena taken into account? Which ones were left out? Were all texts treated equally? Yet, similar questions may be directed towards the bibliographical material which, as the methodological section of this paper pointed out, is far from being an “objective representation” and is rather quite a comprehensive sample of literary material, gathered according to a certain methodology and by no means free from errors.

Furthermore, critics do not work in vacuum, they engage in literary debates and take stands on issues raised by other critics and they are not mere observers but active participants of the events in question. Hence, it is hard to draw a clear border between descriptive and prescriptive activity. Actually, critics quoted in this volume authored 1.6% of the reviews analysed in this study.

Quantitative approaches allow us to go beyond the critical writings and take a look at different dimensions of literary life, especially at those which usually escape critical scrutiny, like the ‘long tail’ of literary production. The obvious question that arises is which construct of literary life is the ‘right’ one: quantitative analysis or critical synthesis? Are we actually talking about the same literary culture? If literary processes are visible most clearly in secondary literature, which is itself a certain interpretation of literary facts, attempted reconstruction of critical diagnoses may be in fact a reconstruction of a certain aspect of literary life, namely actual critical debates.

The main goal of this study was to corroborate certain claims concerning the processes of literary life and the results show that certain type of generalisation is possible. This is why the central, yet problematic variable in this study is time. As we all know too well, the fact that a piece was published in a given year and not the previous or the following one is often a matter of chance. One can blur the time boundaries by using a moving average but this in turn would soften the impact of individual literary events, as the case of Szymborska’s Nobel prize shows. This study marks a point of departure for the further work with *PBL* data, which—freed from the constraints imposed by the goal of the present study—would look at further avenues of macroanalytical research into literary with the use of bibliographic data.

Acknowledgements Research presented in this study was made possible by the National Programme for the Development of Humanities (Research grant: The Polish Literary Bibliography—a knowledge lab on contemporary Polish culture, NR 0061/NPRH3/H11/82/2014), Polish-American Fulbright Commission (Fulbright Senior Award: Literary transition as viewed from above. Quantitative history of Polish literature (1989–2000). Research visit at Stanford University), and COST Association (COST Action CA16213: New Exploratory Phase in Research on East European Cultures of Dissent). The author is greatly indebted to colleagues from the Stanford Literary Lab (Mark Algee-Hewitt, Ryan Heuser, Erik Fredner, J.D. Porter, and Hannah Walser) for their insightful feedback on this project, as well as to the participants of the DFG conference at Villa Vigoni. Special thanks to Cezary Rosiński for his help with the data dump. All mistakes remain my own.

References⁶²

- Allison, Sarah/Heuser, Ryan/Jockers, Matthew L. et al., "Quantitative Formalism: An Experiment," in: *Pamphlets of the Stanford Literary Lab* 1 (2011), <https://litlab.stanford.edu/LiteraryLabPamphlet1.pdf>.
- Bagłajewski, Arkadiusz, "Stan rozproszenia," in: *Odra* 1 (1996), 51–55.
- Bagłajewski, Arkadiusz, *Mapy Dwudziestolecia 1989–2009: Linie Ciągłości*, Lublin 2012.
- Bastian, Mathieu/Heymann, Sebastien/Jacomy, Mathieu, "Gephi: an open source software for exploring and manipulating networks," in: *International AAAI Conference on Weblogs and Social Media* (2009), <https://gephi.org/publications/gephi-bastian-feb09.pdf>.
- Bode, Katherine, *Reading by numbers. Recalibrating the literary field*, London 2014.
- Bode, Katherine, "The Equivalence of 'Close' and 'Distant' Reading. Or, toward a New Object for Data-Rich Literary History," in: *Modern Language Quarterly* 78/1 (2017), 77–106.
- Czapliński, Przemysław/Śliwiński, Piotr, *Literatura Polska 1976–1998: Przewodnik po prozie i poezji*, Kraków 1999.
- Czapliński, Przemysław/Leciński, Maciej/Szybowicz, Eliza et al. (eds.), *Kalendarium Życia Literackiego 1976–2000: Wydarzenia, Dyskusje, Bilanse*, Kraków 2003.
- Czapliński, Przemysław, *Powrót centrali: Literatura w nowej rzeczywistości*, Kraków 2007.
- Czapliński, Przemysław, *Ruchome Marginesy: Szkice O Literaturze Lat 90*, Kraków 2002.
- Czapliński, Przemysław, *Ślady Przełomu*, Kraków 1997.
- Dunin, Kinga, "Normalka," in: Przemysław Czapliński/Piotr Śliwiński (eds.), *Normalność i konflikty: rozważania o literaturze i życiu literackim w nowych czasach*, Poznań 2006.
- Eder, Maciej, "Style-markers in authorship attribution: a cross-language study of authorial fingerprint," in: *Studies in Polish Linguistics* 6 (2011), 99–114.
- Eder, Maciej/Kestemont, Mike/Rybicki, Jan, "Stylometry with R: a suite of tools," in: *Digital Humanities 2013: Conference Abstracts*, University of Nebraska-Lincoln, NE (2013), 487–489.
- Escarpit, Robert. 1966. *The Book Revolution*. Toronto: Harrap.
- Graham, Shawn, *Converting 2-mode networks to 1-mode networks* (2012), <https://electricarchaeology.ca/2012/02/08/converting-2-mode-networks-to-1-mode-networks/>.
- Gawliński, Stanisław, "Wstęp," in: Stanisław Gawliński/Dorota Siwor (eds.), *Dwie dekady nowej (?) literatury 1989–2009*, Kraków 2011.
- Hoover, David L./Culpeper Jonathan/O'Halloran Kieran, *Digital Literary Studies: Corpus Approaches to Poetry, Prose, and Drama*, New York 2014.
- Jannidis, Fotis/Lauer, Gerhard, "Burrows's Delta and Its Use in German Literary History," in: Matt Erlin/Lynne Tatlock (eds.), *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*, New York 2014, 29–54.
- Jockers, Matthew L., *Macroanalysis. Digital Methods & Literary History*, Chicago 2013.
- Long, Hoyt/So, Richard, "Network Science and Literary History," in: *Leonardo* 46/3 (2013).
- Maryl, Maciej/Wciślik, Piotr: "Remediations of Polish Literary Bibliography: Towards a Lossless and Sustainable Retro-Conversion Model for Bibliographical Data," in: *Digital humanities 2016: Conference Abstracts*, Kraków 2016, 621–623, <http://dh2016.adho.org/abstracts/293>.
- Moretti, Franco, *Distant Reading*, London 2013a.
- Moretti, Franco, "'Operationalizing': or, the function of measurement in modern literary theory," in: *Pamphlets of the Stanford Literary Lab* 6 (2013b), <https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf>.
- Nasiłowska, Anna, *Literatura Okresu Przejściowego 1975–1996*, Warszawa 2006.
- Opsahl, Tore, *Structure and Evolution of Weighted Networks*, University of London (Queen Mary College), London 2009, 104–122, unpublished thesis, <http://toreopsahl.com/publications/thesis/>.

⁶²All digital references were last viewed on April 20, 2018.

- Opsahl, Tore/Agneessens, Filip/Skvoretz, John, "Node centrality in weighted networks: Generalizing degree and shortest paths," in: *Social Networks* 32/3 (2010), 245–251.
- Ramsay, Stephen, *Reading Machines. Towards an Algorithmic Criticism*, Urbana 2011.
- Rybicki, Jan, "Pierwszy rzut oka na stylometryczną mapę literatury polskiej," in: *Teksty Drugie* 2 (2014), 106–128.
- Shawn, Graham/Milligan, Ian/Weingart, Scott, *Exploring Big Historical Data: The Historian's Macroscope*, London 2015.
- So, Richard/Long, Hoyt, "Network analysis and the sociology of Modernism," in: *Boundary* 2 40/2 (2013), 147–82.
- Schöch, Christof/Pielström, Steffen, "Für eine computergestützte literarische Gattungsstilistik," in: *1. Jahrestagung der Digital Humanities im deutschsprachigen Raum (DHD)* (2014), http://dig-hum.de/sites/dig-hum.de/files/Schoch-Pielstrom_2014_Gattungsstilistik.pdf.
- Sławiński, Janusz, "Zanik centrali," in: *Kresy* 18 (1994), 14–16.
- Umerle, Tomasz, "Rethinking the Potential of Documentation of Culture as a Data Gathering Practice," in: *Proceedings from the Document Academy* 4 (2017), DOI <https://doi.org/10.35492/docam/4/2/15>.

Online Resources

- Science of Science (Sci2) Tool. Indiana University and SciTech Strategies*, <https://sci2.cns.iu.edu>.
- Polska Bibliografia Literacka*, <https://pbl.ibl.waw.pl/>.
- Signorell, Andri/Aho, Ken/Alfons, Andreas et al., *DescTools: Tools for Descriptive Statistics, R package version 0.99.24* (2018), <https://cran.r-project.org/package=DescTools> (retrieved on April 20, 2018).
- Zeileis, Achim, *ineq: Measuring Inequality, Concentration, and Poverty, R package version 0.2–13* (2014), <https://CRAN.R-project.org/package=ineq>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Quantitative Semantik. *Word Embedding Models* für literaturwissenschaftliche Fragestellungen

Christof Schöch

1 Einleitung

Zum Verhältnis der hermeneutischen und der quantitativen Literaturwissenschaften besteht eine Reihe von Argumentationsmustern, nach denen die letzteren von Fall zu Fall als Retter, Herausforderer, Sargnagel oder Ergänzung der ersteren gesehen werden. Alle diese Argumentationsmuster haben ihren Charme und ihre Schattenseiten, ihre Berechtigung und ihre Grenzen. In der Tat entscheidet sich die Frage nach dem Verhältnis von hermeneutischen und quantitativen Verfahren in den Literaturwissenschaften nicht ein für alle Mal. Ihre Beantwortung hängt jeweils davon ab, welche Gruppe von Methoden und Verfahren aus dem reichen Fundus der quantitativen Literaturwissenschaften gerade infrage steht und auf welchem Stand der Entwicklung sich diese befindet.

Der vorliegende Beitrag behandelt einen spezifischen Bereich der Literaturwissenschaften: den der Zuschreibung von Bedeutung zu Wörtern und der Interpretation von literarischen Texten. Im Gegensatz zu anderen Teilbereichen der Literaturwissenschaften, in denen digitale Methoden und Verfahren mittlerweile als gut etablierte, wenn nicht gar dominante Teile des Methodenrepertoires verstanden werden können – man denke an die Editionswissenschaften (die sich mit der Repräsentation textueller Phänomene befassen) oder die Autorschaftsattribuion (die sich mit der lexikalischen und stilistischen Ähnlichkeit von Texten befasst) –, ist die Interpretation von Texten ein Bereich, der die

Förderhinweis Die Arbeit an diesem Beitrag erfolgte im Rahmen der Nachwuchsgruppe *Computergestützte literarische Gattungsstilistik*, die vom *Bundesministerium für Forschung und Bildung (BMBF)* unter FKZ 01UG1508 gefördert wird.

C. Schöch (✉)
Universität Trier, Trier, Deutschland
E-Mail: schoech@uni-trier.de

Herausforderung durch quantitative Verfahren lange Zeit erfolgreich abwehren konnte. In diesem Bereich schien bisher nur ein arbeitsteiliges Verhältnis denkbar, bei dem digitale Methoden und Tools die Suche in Texten und die Exploration von Texten erleichtern, aber der Beitrag der Forschenden unabdingbar bleibt, wenn es darum geht, Texten Bedeutung zuzuschreiben und kontextabhängige Interpretationen anzubieten.

Auch wenn die Geschichte der quantitativen Semantik weit zurückreicht,¹ so haben sich doch in den letzten fünf bis zehn Jahren neue statistische Zugänge zu Bedeutung in Texten in einer äußerst dynamischen Weise entwickelt. Dazu gehören unter anderem das *Topic Modeling*² und die *Sentiment-Analyse*.³ Außerdem gilt dies für den großen Bereich der Vektorsemantik,⁴ in den auch die hier verhandelten *Word Embedding Models* (im Folgenden *WEM*) fallen. Für den teils spektakulären Aufschwung dieser Methoden sind mehrere Entwicklungen entscheidend gewesen: erstens massive Digitalisierungsaktivitäten und das ubiquitär gewordene Internet, durch die mehr und mehr umfangreiche Textdaten entstehen und verfügbar sind,⁵ zweitens die Tatsache, dass auch die verfügbare Rechenleistung massiv angestiegen ist und insbesondere die parallele Ausführung von

¹ Siehe hierzu bspw. Michael Stubbs, „Notes on the History of Corpus Linguistics and Empirical Semantics“, in: Marja Nenonen/Sinikka Niemi (Hg.), *Collocations and Idioms*, Joensuu, Finnland, 2007, 317–329.

² David M. Blei/Andrew Y. Ng/Michael I. Jordan, „Latent Dirichlet Allocation“, in: *Journal of Machine Learning Research* 3 (2003), 993–1022, <http://dl.acm.org/citation.cfm?id=944919.944937>; Mark Steyvers/Tom Griffiths, „Probabilistic Topic Models“, in: Thomas K. Landauer/Danielle S. McNamara/Simon Dennis u. a. (Hg.), *Latent Semantic Analysis. A Road to Meaning*, Mahwah, NJ, 2006; David M. Blei, „Probabilistic Topic Models“, in: *Communications of the ACM* 55/4 (2012), 77–84, <https://doi.org/10.1145/2133806.2133826>; literaturwissenschaftliche Anwendungsbeispiele sind u. a. Cameron Blevins, „Topic Modeling Martha Ballard’s Diary“, in: *Historying* (April 2010), <http://historying.org/2010/04/01/topic-modeling-martha-ballards-diary/> oder Christof Schöch, „Topic Modeling Genre. An Exploration of French Classical and Enlightenment Drama“, in: *Digital Humanities Quarterly* 11/2 (2017), <http://www.digitallhumanities.org/dhq/vol/11/2/000291/000291.html> (letzter Aufruf 8.5.2018).

³ Bo Pang/Lillian Lee, „Opinion Mining and Sentiment Analysis“, in: *Found. Trends Inf. Retr.* 2/1–2 (2008), 1–135, <https://doi.org/10.1561/1500000011>; für literaturwissenschaftliche Anwendungen siehe Saif Mohammad, „From Once Upon a Time to Happily Ever After. Tracking Emotions in Novels and Fairy Tales“, in: *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, Stroudsburg, PA, 2011, 105–114, <http://dl.acm.org/citation.cfm?id=2107636.2107650> (letzter Aufruf 8.5.2018), oder Fotis Jannidis/Isabella Reger/Albin Zehe, „Analyzing Features for the Detection of Happy Endings in German Novels“, in: *arXiv:1611.09028 [cs.IR]* (28. November 2016), <http://arxiv.org/abs/1611.09028>.

⁴ Siehe einführend Dominic Widdows, *Geometry and Meaning*, Stanford, CA, 2004.

⁵ Man denke an Textsammlungen wie *ECCO* (*Eighteenth-Century Collection Online*; siehe <https://quod.lib.umich.edu/e/ecco/>), die *Digitale Bibliothek* von *TextGrid* (siehe <https://textgrid.de/digitale-bibliothek/>), die Sammlungen des *Hathi Trust* (siehe <https://www.hathitrust.org/>) oder das *Google-Ngram*-Korpus (siehe <http://storage.googleapis.com/books/ngrams/books/datasetsv2.html> [jeweils letzter Aufruf 8.5.2018]).

Matrizenoperationen, wie sie für massiv iterative Verfahren wie neuronale Netze wichtig sind, durch die Verfügbarkeit darauf spezialisierter, leistungsfähiger Grafikkarten auch ohne aufwendige Cluster-Architekturen möglich geworden ist. Und drittens das ausgeprägte Interesse nicht nur der akademischen Forschung, sondern auch der datenorientierten digitalen Großkonzerne, die in die Entwicklung entsprechender Methoden investieren. Gleichzeitig sind neue theoretische Einsichten in Modelle der mentalen Repräsentation konzeptueller Bedeutung⁶ und Arbeiten zur Beziehung zwischen der schwer fassbaren lexikalischen Bedeutung und der mathematisch gut greifbaren Geometrie erschienen.⁷ Zudem haben die neuen semantischen Technologien in Anwendungsbereichen wie der kognitiven Psychologie (u. a. bspw. *Language Acquisition*) oder dem *Natural Language Processing* (*Machine Translation*, Chat-Systeme) bereits erstaunliche Fortschritte ermöglicht. Aus diesem Grund stellt sich die Frage derzeit erneut: Sind quantitative Ansätze zu textueller Semantik im Begriff, die Rolle des Herausforderers etablierter Methoden der Textinterpretation in den Literaturwissenschaften einzunehmen?

Es ist nicht möglich, im Rahmen eines Einzelbeitrags alle Methoden der quantitativen Semantik zu diskutieren, die für diese Fragestellung relevant wären. Der vorliegende Beitrag konzentriert sich daher auf eine derzeit besonders aktuelle Variante der Vektorsemantik, die *WEM* oder auch *Distributional Semantic Models*, die in verschiedenen Varianten und Implementierungen vorliegen. Während die klassische Vektorsemantik auf einer Term-Dokument-Matrix operiert und in erster Linie Aufgaben wie das *Retrieval*, die Klassifikation oder das *Clustering* von Dokumenten ermöglicht, beziehen sich *WEM* auf eine Term-Kontext-Matrix, wobei der Kontext nicht ein ganzes Dokument, sondern ein bestimmtes Fenster um das Zielwort herum ist. Sie erlauben es entsprechend, die Ähnlichkeit von Einzelwörtern zu ermitteln und darauf aufbauende Tasks wie Klassifikation oder Clustering, die Disambiguierung von Wortbedeutungen, die Bildung von Thesauri oder *Query Expansion* zu ermöglichen.⁸

Die Ziele des vorliegenden Beitrags sind, einige Grundprinzipien von *WEM* darzustellen, einen orientierenden Überblick über aktuelle Anwendungsszenarien in den digitalen Literaturwissenschaften anzubieten und über die Implikationen dieser Methoden für die Literaturwissenschaften der Zukunft nachzudenken. Der Beitrag wird daher, erstens, die konzeptionellen und statistischen Grundlagen von *WEM* skizzieren. Zweitens werden einige grundlegenden Abfragemechanismen solcher Modelle aufgezeigt. Drittens werden einige Anwendungsbereiche von *WEM* im Dienste anderer quantitativer Verfahren der Textanalyse erläutert. Schließlich werden mehrere eigenständige Anwendungsbereiche von *WEM* für Fragestellungen aus den Literaturwissenschaften dargestellt, bevor ein Fazit erfolgt.

⁶Z. B. Peter Gardenfors, *Conceptual Spaces. The Geometry of Thought*, Cambridge, MA, 1998.

⁷Z. B. Widdows (Anm. 4).

⁸Einen Überblick bieten Peter T. Turney/Patrick Pantel, „From Frequency to Meaning. Vector Space Models of Semantics“, in: *Journal of Artificial Intelligence Research* 37 (2010), 141–188, <https://arxiv.org/abs/1003.1141> (letzter Aufruf 8.5.2018).

2 Was sind *Word Embedding Models (WEM)*?

Die Annahme, die der Vektorenssemantik allgemein ebenso wie den *Word Embedding Models* zugrunde liegt, ist die sogenannte *Distributional Hypothesis*, der zufolge Wörter, die in ähnlichen Kontexten erscheinen, auch eine ähnliche Bedeutung (und ggf. ähnliche morpho-syntaktische Funktion) haben. Diese Annahme steht im Gegensatz zu formalen und referenziellen Theorien der Wortsemantik, die annehmen, dass die Bedeutung eines Wortes von seinem abstrakten oder konkreten Referenten abhängt. Dieses Prinzip der distributionellen Semantik wurde schon in den 1950er Jahren formuliert⁹ und ihre konkrete Nutzung für die kognitive Psychologie und quantitative Semantik mindestens seit den späten 1980er Jahren exploriert.¹⁰ Zu den wichtigsten frühen methodischen Ansätzen zählen hier die Arbeiten von Gerald Salton u. a. (1975) zum *Vector Space Model* für die Repräsentationen von Dokumenten und Queries im *Information Retrieval*,¹¹ von Rumelhart und McClelland (1986) zum *Parallel Distributed Processing*¹² sowie die von Deerwester u. a. 1990 entwickelte *Latent Semantic Analysis (LSA)*.¹³ Letztere beruht auf einer Methode aus der Matrizenrechnung, der sogenannten *Singular Value Decomposition (SVD)*, um Kookkurrenzen von Wörtern zu berechnen.¹⁴ Landauer u. a. schreiben:

Latent Semantic Analysis (LSA) is a theory and method for extracting and representing the contextual-usage meaning of words by statistical computations applied to a large corpus of text (Landauer and Dumais, 1997). The underlying idea is that the aggregate of all the word contexts in which a given word does and does not appear provides a set of mutual constraints that largely determines the similarity of meaning of words and sets

⁹ Siehe Zellig Harris, „Distributional Structure“, *Word* 10/2–3 (1954), 146–62, <https://doi.org/10.1080/00437956.1954.11659520>; John R. Firth, „A Synopsis of Linguistic Theory, 1930–1955“, in: *Studies in Linguistic Analysis* (1957), 1–32.

¹⁰ Zwei Pioniere der Anwendung der distributionellen Hypothese in der kognitiven Psychologie formulieren: „The basic idea is simply that words with similar meanings will tend to occur in similar contexts, and hence word co-occurrence statistics can provide a natural basis for semantic representations“, siehe John A. Bullinaria/Joseph P. Levy, „Extracting Semantic Representations from Word Co-Occurrence Statistics. A Computational Study“, in: *Behavior Research Methods* 39/3 (2007), 510–526, <https://doi.org/10.3758/BF03193020>.

¹¹ Gerard Salton/Andrew Wong/Chung-Shu Yang, „A Vector Space Model for Automatic Indexing“, in: *Communications of ACM* 18/11 (1975), 613–20, <https://doi.org/10.1145/361219.361220>.

¹² James McClelland/David E. Rumelhart (Hg.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Cambridge, MA, 1986.

¹³ Scott Deerwester/Susan T. Dumais/George W. Furnas u. a., „Indexing by Latent Semantic Analysis“, in: *Journal of the American Society for Information Science* 41/6 (1990), 391–407, [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<391::AID-AS11>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-AS11>3.0.CO;2-9).

¹⁴ Ebd.

of words to each other. The adequacy of LSA's reflection of human knowledge has been established in a variety of ways.¹⁵

Frühe Anwendungsszenarien der *LSA* und ähnlicher Verfahren kamen insbesondere aus der kognitiven Psychologie¹⁶ und dem *Information Retrieval*. Heute ist *LSA* nicht nur als eine wegweisende Methode der Vektorsemantik zu verstehen, sondern auch als Vorläufer des äußerst populären *Topic Modelings* (bzw. einer seiner Varianten, der *Latent Dirichlet Allocation* oder *LDA*).

Die *WEM* erfuhren 2013 einen weiteren, entscheidenden technischen Durchbruch mit einem Vorschlag von Mikolov u. a.,¹⁷ wie *WEM* auch auf der Grundlage sehr umfangreicher Textdatensätze effizient ermittelt werden können: das sogenannte *word2vec*-Verfahren zur Transformation von Wörtern zu Vektoren. Aufbauend auf Erfahrungen mit verschiedenen Varianten neuronaler Netze, schlagen Mikolov u. a. zwei Verfahren vor, die beide neuronale Netze einsetzen, dabei aber Modellarchitekturen verwenden, die durch einige vereinfachende Annahmen weniger komplex und rechenintensiv als vorige Architekturen sind.¹⁸ Zugleich, so das Ziel von Mikolov u. a., sollte die Qualität der resultierenden *WEM* nicht nur gleich bleiben, sondern auch noch gesteigert werden können, zumindest unter der Voraussetzung, dass auch entsprechend umfangreiche Textdaten für das Training vorhanden sind. Diese beiden Architekturen nennen die

¹⁵Thomas K. Landauer/Peter W. Foltz/Darrell Laham, „Introduction to Latent Semantic Analysis“, in: *Discourse Processes* 25 (1998), 259–284, hier: 259. Siehe auch: Thomas K. Landauer/Susan T. Dumais, „A solution to Plato's problem. The latent semantic analysis theory of acquisition, induction, and representation of knowledge“, in: *Psychological Review* 104/2 (1997), 211–240, <https://doi.org/10.1037/0033-295X.104.2.211>.

¹⁶Für einen Überblick siehe Bullinaria/Levy (Anm. 10).

¹⁷Tomas Mikolov/Kai Chen/Greg Corrado u. a., „Efficient Estimation of Word Representations in Vector Space“, in: *arXiv:1301.3781 [cs.CL]* (7. September 2013), <http://arxiv.org/abs/1301.3781> (letzter Aufruf 8.5.2018).

¹⁸Dabei gehört *LSA* zu den „count-based methods“ in der Vektorsemantik, während *word2vec* zu den „predictive models“ gehört. Siehe hierzu: „Count-based methods compute the statistics of how often some word co-occurs with its neighbor words in a large text corpus, and then map these count-statistics down to a small, dense vector for each word. Predictive models directly try to predict a word from its neighbors in terms of learned small, dense *embedding vectors* (considered parameters of the model).“ („Vector Representations of Words“, in: *Tensor Flow Tutorials* (2018), <https://www.tensorflow.org/tutorials/word2vec> (letzter Aufruf 8.5.2018). Siehe zu dieser Unterscheidung auch Marco Baroni/Georgiana Dinu/Germán Kruszewski, „Don't Count, Predict! A Systematic Comparison of Context-counting vs. Context-predicting Semantic Vectors“, in: *Proceedings of Association for Computational Linguistics (ACL)*, Bd. 1, Baltimore, MD, 2014, 238–247, <http://acl2014.org/acl2014/P14-1/pdf/P14-1023.pdf> (letzter Aufruf 8.5.2018).

Tab. 1 Kontextwörter und Zielwort in der CBOW-Architektur

Kontextwörter als Input	Zielwort
„Morellet hatte ein ... unter den Dächern“	„Zimmer“
„hatte ein Zimmer ... den Dächern im“	„unter“
„ein Zimmer unter ... Dächern im achten“	„den“
„Zimmer unter den ... im achten Stock“	„Dächern“

Autoren *CBOW* (Continuous Bag-of-Words) und *Skip-Gram*. Weitere derzeit viel genutzte Architekturen sind u. a. *GloVe*¹⁹, *FastText*²⁰ und *SVD-PPMI*²¹.

Stellvertretend für die genannten Verfahren soll hier das *CBOW*-Verfahren in seinen Grundzügen vorgestellt werden. Wie auch bei dem *Skip-Gram*-Verfahren wird hier ein einfaches neuronales Netzwerk mit drei Schichten eingesetzt: einem *Input Layer*, einem *Projektionslayer* und einem *Output Layer*. Das Netzwerk wird trainiert, indem nach und nach Trainingsdaten präsentiert werden, wobei ein Satzfragment nach dem anderen als Input dient. Aufgabe des Netzwerkes ist es im Falle der *CBOW*-Architektur, auf der Grundlage von jedem der Kontextwörter (in einem Fenster von beispielsweise drei Wörtern vor und nach dem Zielwort) jeweils das Zielwort vorherzusagen. Das Netzwerk bekommt also Satzfragmente präsentiert, in denen das zentrale Zielwort fehlt. Ein Satz aus Georges Perecs *La Vie mode d'emploi* (*Das Leben Gebrauchsanweisung*) lautet beispielsweise: „Morellet hatte ein Zimmer unter den Dächern, im achten Stock.“

Aus diesem Satz lassen sich unter anderem die Inputfragmente in Tab. 1 ableiten, wobei das jeweilige Zielwort in der Mitte der Sequenz ausgelassen wird.

Ein wichtiger Vorteil einer solchen Architektur ist, dass die Trainingsdaten (Kontextwörter als Input und korrekte Antwort) unmittelbar und automatisch aus den Textdaten abgeleitet werden können, ohne dass eine manuelle Annotation der Zielwörter notwendig ist. So kann, das Vorhandensein ausreichender Mengen an Text vorausgesetzt, leicht das notwendige, sehr umfangreiche Trainingsmaterial eingesetzt werden.

Der *Input Layer* repräsentiert nun nach und nach jeweils die Kontextwörter, wobei die Repräsentation nach dem Prinzip eines *One-Hot Vectors* geschieht, d. h. als ein Vektor, in dem jede Position einem Wort im Gesamtvokabular entspricht, und das Wort kodiert wird, indem nur die eine ihm entsprechende Position im Vektor den Wert 1 bekommt, während alle anderen 0 erhalten. Der *Output Layer*

¹⁹ Jeffrey Pennington/Richard Socher/Christopher D. Manning, *Glove. Global Vectors for Word Representation*, Stanford 2014, <http://citeseerx.ist.psu.edu/viewdoc/citations;jsessionid=B90254BA67F435112ACC1AC456222FA9?doi=10.1.1.671.1743> (letzter Aufruf 8.5.2018).

²⁰ Armand Joulin/Edouard Grave/Piotr Bojanowski u. a., „Bag of Tricks for Efficient Text Classification“, in: *arXiv:1607.01759* (Juli 2016), <http://arxiv.org/abs/1607.01759>.

²¹ Omer Levy/Yoav Goldberg/Ido Dagan, „Improving Distributional Similarity with Lessons Learned from Word Embeddings“, in: *Transactions of the Association for Computational Linguistics* 3 (2015), 211–225, <https://www.transacl.org/ojs/index.php/tacl/article/view/570> (letzter Aufruf 8.5.2018).

kodiert die Vorhersage des Zielworts durch das Netzwerk, indem jedes Wort im Vektor des Gesamtvokabulars einen Wert zwischen 0 und 1 zugeordnet bekommt, der seiner Wahrscheinlichkeit im gegebenen Kontext entspricht. Im Falle des Inputs „Morellet hatte ein ... unter den Dächern“ werden mehrere Wörter ähnliche Wahrscheinlichkeiten haben, denn es könnte sich ja hier aufgrund des inhaltlichen und grammatikalischen Kontextes nicht nur um ein Zimmer (wie hier tatsächlich der Fall), sondern beispielsweise auch um ein Appartement oder ein Labor handeln. Weil das Netzwerk sehr viele solcher und ähnlicher Satzfragmente präsentiert bekommt, lernt es aber sowohl, welches Wort für jeden Kontext die wahrscheinlichste Variante ist, als auch, welche Wörter in ähnlichen Kontexten ähnliche Wahrscheinlichkeiten haben.

Zugleich entsteht in dem *Projection Layer* eine Repräsentation der Zielwörter. (Diese ist wesentlich wichtiger als der eigentlich gelernte Task, der, für sich genommen, nur wenige konkrete Anwendungsszenarien unterstützt, wie beispielsweise die automatische Korrektur von durch OCR ermitteltem Text.) Der *Projection Layer* vermittelt zwischen *Input Layer* und *Output Layer* und repräsentiert jedes Wort des Vokabulars in einem gemeinsamen Vektorraum. Sie wird mit zufälligen Werten initialisiert und dann durch das Zusammenspiel einer Optimierungsfunktion mit den Trainingsdaten iterativ verbessert. Wörter, die in ähnlichen Kontexten vorkommen können, müssen sich morpho-syntaktisch und semantisch ähnlich sein, weil sie in ähnliche Leerstellen ‚passen‘ müssen. Aufgrund der Anforderung, dass sie in ähnlichen Kontexten ähnliche Wahrscheinlichkeiten haben, ähnelt sich schließlich auch ihre Repräsentation in der *Projection Layer*. Anders als *Input Layer* und *Output Layer*, deren Dimensionalität dem Umfang des Vokabulars (also der Anzahl der unterschiedlichen Wörter) in den Trainingsdaten entspricht, hat der *Projection Layer* in der Regel eine weit geringere Dimensionalität. Üblich ist für die Anzahl der Dimensionen eines WEM-Modells zwischen 100 und mehreren Hundert. Es ergibt sich also für den *Projection Layer* eine Matrix der Form ‚Vokabular x Dimensionen‘. Gegenüber den hochdimensionalen, überwiegend mit ‚0‘ besetzten Kontextvektoren, die daher auch als *Sparse Representation* bezeichnet werden, ist diese *Projection Layer* ein Vektorraum, in dem jedes Wort des Vokabulars einen je unterschiedlichen Wert in jeder der vergleichsweise wenigen Dimensionen hat. Daher spricht man auch von einer *Dense Vector Representation*.

Der wesentliche Unterschied zwischen der CBOW-Architektur und der ebenfalls von Mikolov u. a. vorgeschlagenen *Skip-Gram*-Architektur ist folgender: In der CBOW-Architektur werden Kontextwörter präsentiert, und es sollen Wahrscheinlichkeiten für die Leerstelle in ihrem Zentrum vorhergesagt werden, während bei der *Skip-Gram*-Architektur umgekehrt ein Zielwort präsentiert wird und die Wahrscheinlichkeiten der Kontextwörter vorhergesagt werden sollen. Das Ergebnis, der *Projection Layer*, ist aber vergleichbar. Beide Verfahren haben etwas unterschiedliche Eigenschaften, was die Sensitivität für morphologische, syntaktische und semantische Ähnlichkeit von Wörtern angeht und was die Qualität der resultierenden *Word Embeddings* für verschiedene Aufgaben wie semantische Ähnlichkeit, Analogiebildungs-Tasks oder morphologische

Ableitungs-Tasks betrifft. Und auch die jeweils verwendeten Optimierungsverfahren und andere Parameter (wie bspw. die Größe des Kontextfensters, die Anzahl der Dimensionen des *Projection Layers* und natürlich die Menge des Trainingsmaterials) können sich auf die spezifischen Eigenschaften eines resultierenden *WEM* auswirken.²² Generell kann man dennoch festhalten, dass man in einem solchen Modell sowohl den Grad der räumlichen Nähe als auch die räumlichen Beziehungen zwischen Wörtern nutzen kann, um Informationen über die semantische Ähnlichkeit oder die semantischen und morpho-syntaktischen Beziehungen zwischen Wörtern zu ermitteln. Dabei wird die semantische Ähnlichkeit zweier Wörter in der Regel als die Cosinus-Distanz zwischen den beiden Vektoren der Wörter im *WEM* gemessen,²³ auch wenn selbstverständlich auch andere Distanz- bzw. Ähnlichkeitsmaße denkbar sind.

Dabei ist das Verfahren nicht prinzipiell auf die Modellierung von Einzelwörtern begrenzt. Einige interessante Erweiterungen des Verfahrens betreffen vielmehr die Modellierung nicht nur von Einzelwörtern, sondern auch von Wortverbindungen („Los Angeles“, „New York Times“), Phrasen und idiomatischen Ausdrücken („take it easy“, „bite the dust“) oder auch von größeren Einheiten bzw. umgekehrt die Berücksichtigung kleinerer Einheiten als Wörter (bspw. Morpheme). Im Folgenden werden jedoch nur Modelle auf der Ebene von Einzelwörtern eine Rolle spielen.

3 Grundlegende Abfragemöglichkeiten von *WEM*

Die folgende Darstellung²⁴ einiger Einsatzmöglichkeiten von *WEM* greift auf zwei Modelle zurück: Das eine wurde auf der Grundlage von 1200 französischen Romanen (erschieden zwischen 1900 und 2010) erstellt, das andere auf der Grundlage der vollständigen französischen *Wikipedia*.²⁵ Um den Einfluss der Morphosyntax zumindest zu reduzieren, sind die beiden hier verwendeten Modelle einerseits mit lemmatisierten Wortformen berechnet worden; andererseits ist die Information über die Wortart jedem Lemma mitgegeben worden. (Für einige Details zu den verwendeten Texten und den wesentlichen Parametern der Modelle siehe die Hinweise im Anhang.)

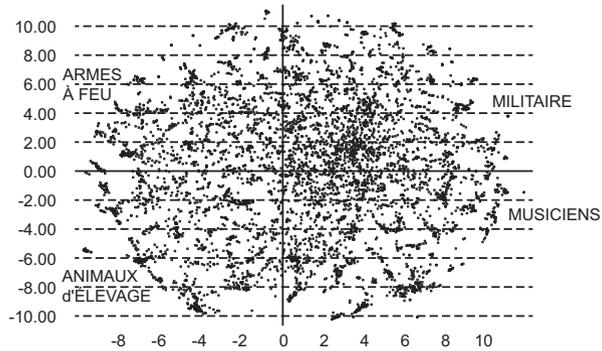
²²All diese Aspekte ebenso wie die verschiedenen Anwendungsszenarien sind Gegenstand intensiver aktueller Forschung im *Natural Language Processing* und der Computerlinguistik, bspw. bei den aktuellen Jahreskonferenzen der *ACL (Association for Computational Linguistics)* oder bei der *EMNLP (Empirical Methods in Natural Language Processing)*.

²³Siehe Turney/Pantel (Anm. 8).

²⁴Hinweis: Diese und einige der folgenden Abbildungen sind in interaktiver Form online verfügbar unter: <https://github.com/cligs/projects/tree/master/2017/vigoni> (letzter Aufruf 8.5.2018). Alle Abbildungen wurden von Christof Schöch erstellt und sind unter der Lizenz *Creative Commons Attribution International 4.0* nachnutzbar.

²⁵Stand Juni 2017.

Abb. 1 Projektion eines Teils eines *Word Embedding Models* auf zwei Dimensionen²⁶



Es ist nicht trivial, sich einen Eindruck von einem *Word Embedding Model* zu machen, in dem jedes Wort aus einem Vektor mehrerer Hundert Dimensionen besteht. Ein mögliches Vorgehen ist, bestimmte Teilmengen des Gesamtvokabulars zu selektieren und diese in zwei Dimensionen zu visualisieren. Abb. 1 zeigt eine solche Darstellung, wobei hier die 5800 häufigsten Substantive im *Wikipedia*-Modell von den 300 Dimensionen des Modells auf zwei Dimensionen projiziert wurden, wofür eine spezifische Methode der Dimensionalitätsreduktion eingesetzt wurde (*t-SNE*).²⁷ Es sind hier Gruppen oder Cluster von Wörtern erkennbar, und bei näherer Betrachtung zeigt sich, dass diese Gruppen jeweils aus semantisch verwandten Wörtern bestehen. Einige dieser Gruppen wurden in der Abbildung manuell mit einem Label versehen. Das Label ‚music‘ (rechts unten) benennt beispielsweise eine Gruppe, in der die folgenden Wörter vorkommen: „bassiste“, „trompettiste“, „instrumentaliste“, „percussionniste“, „répertoire“, „chanteur“, „improvisation“, „musicologue“, „musique“, „fanfare“ und einige mehr.

Aufgrund des schieren Umfangs eines solchen *WEM*, sowohl was die Anzahl der Wörter als auch die Anzahl der Dimensionen angeht, bleibt ein solcher Ansatz aber unbefriedigend. Zudem werden so nur die Ähnlichkeitsbeziehungen im zweidimensionalen Raum sichtbar, der ja aufgrund einer massiven Dimensionalitätsreduktion und damit unter Informationsverlust entstanden ist, nicht aber die eigentlich im vieldimensionalen Raum gegebenen Näheverhältnisse. Einen besseren Eindruck von einem *WEM* bekommt man, indem man verschiedene, grundlegende Abfragen vornimmt. Es folgt daher nun die Illustration einiger solcher Abfragemöglichkeiten, bevor auf weiterführende und inhaltlich komplexere Anwendungsbereiche für Fragestellungen aus den Digitalen Literaturwissenschaften eingegangen wird.

²⁶ 5800 häufigste Substantive, *Wikipedia*-Modell. Bildlizenz: *Creative Commons Attribution International 4.0 (CC-BY)*.

²⁷ Vgl. Laurens van der Maaten/Geoffrey Hinton, „Visualizing Data Using T-SNE“, in: *Journal of Machine Learning Research* 9 (2008), 2579–2605, <http://www.jmlr.org/papers/v9/vandermaaten08a.html> (letzter Aufruf 8.5.2018).

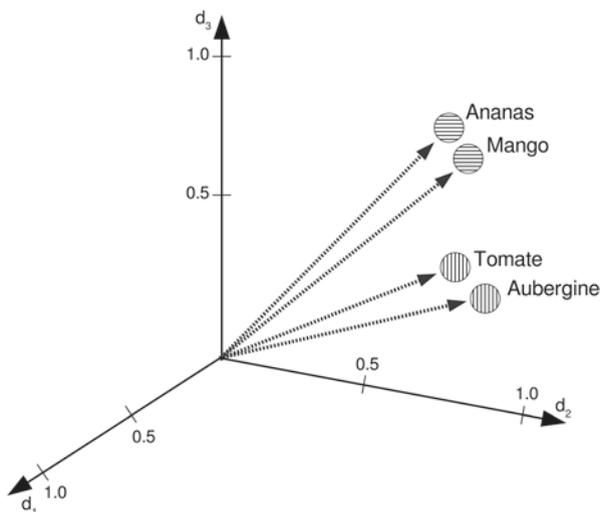


Abb. 2 Räumliche Nähe und semantische Ähnlichkeit im Vektorraum (illustratives, künstliches Beispiel)²⁸

3.1 Ähnlichkeit von Wörtern

Ein *WEM* kodiert, wie bereits dargelegt, die räumlichen Beziehungen von Wörtern in einem vieldimensionalen Vektorraum. Die übliche Interpretation der räumlichen Nähe ist dabei semantische Ähnlichkeit, wobei berücksichtigt werden sollte, dass auch morpho-syntaktische Ähnlichkeit eine wichtige Rolle spielen kann.

Abb. 2 zeigt eine abstrakte, vereinfachte Repräsentation von Wörtern in einem solchen Vektorraum. Die Abbildung illustriert die Idee, dass sich Wörter mit besonders ähnlichen Bedeutungen besonders nahe stehen werden: *Ananas* und *Mango* befinden sich nahe beieinander, weil es sich jeweils um essbares Obst handelt; das Gleiche gilt für *Aubergine* und *Tomate*, weil es sich jeweils um Gemüse handelt. Während sich also verschiedene Obstsorten sehr nahestehen und verschiedene Gemüse ebenfalls, befindet sich zwischen der Gruppe der Obstsorten und der Gruppe der Gemüse eine etwas größere Distanz. Zugleich teilen sich die Obstsorten und Gemüse auch wiederum einen gemeinsamen Bereich des Vektorraums, denn es handelt sich ja in allen Fällen um Lebensmittel. In einem wirklichen *WEM* hat ein solcher Vektorraum wie erwähnt nicht wie hier nur drei, sondern mindestens 50 bis mehrere Hundert Dimensionen. Und natürlich sind nicht nur vier Wörter in ihm verortet, sondern der größte Teil des in einer Textsammlung vorfindbaren Vokabulars, das leicht in einer Größenordnung von mehreren zehntausend Wörtern liegen kann.

²⁸ Bildlizenz: *Creative Commons Attribution International 4.0 (CC-BY)*.

Das Prinzip, dass räumliche Nähe als Indikator für semantische Ähnlichkeit gilt, kann zunächst einmal dafür eingesetzt werden, um die einem bestimmten Suchbegriff nächsten – also semantisch ähnlichsten – Begriffe zu finden. Dies wird nun an realen Suchabfragen in dem *WEM* gezeigt, das auf der französischen *Wikipedia* basiert. Der folgende Output ergibt sich bei Verwendung des *Wikipedia*-Modells, wenn man nach den zehn Wörtern fragt, die dem Substantiv „poésie“ (dt. Dichtung) am ähnlichsten sind:

Abfrage: „[poésie_nom, 10]“

Ergebnis: {"poétique_adj": 0.841208, "poème_nom": 0.789321, "prose_nom": 0.732649, "littérature_nom": 0.714615, "poète_nom": 0.703788, "poétique_nom": 0.701020, "poésie_nam": 0.700259, "anthologie_nom": 0.695312, "littéraire_adj": 0.654449, "sonnet_nom": 0.650843}

Die Werte liegen grundsätzlich zwischen -1 (geringste Nähe bzw. geringste semantische Ähnlichkeit) und 1 (größte Nähe bzw. größte semantische Ähnlichkeit). Wir sehen hier, dass „poétique“ und „poème“ dem Suchwort „poésie“ dem Modell zufolge semantisch am ähnlichsten sind, was natürlich ebenso einleuchtend wie trivial ist. Interessant, und ein erster Hinweis auch auf mögliche Komplikationen mit einem solchen Modell, ist die Tatsache, dass auch „prose“ einen sehr hohen Ähnlichkeitswert zu „poétique“ hat. Das mag daran liegen, dass der Ausdruck „prose poétique“ sehr häufig in der *Wikipedia* vorkommt oder dass die beiden Begriffe schlicht oft gemeinsam verwendet werden; es wird jedoch deutlich, dass hier die Tatsache, dass „prose“ und „poésie“ in vielen Kontexten als Gegensätze verstanden werden, nicht eindeutig ablesbar wird. Im Kontext einer thematisch äußerst breit gefassten französischen Enzyklopädie sind Prosa und Poesie sich semantisch sehr nahe, im spezifischen Kontext des Redens über literarische Texte mag das anders sein. Jedes *WEM*, so zeigt sich hier bereits, spiegelt die Verhältnisse wider, die im jeweils zugrundeliegenden Textmaterial vorliegen.

Eine weitere Abfragemöglichkeit, die ebenfalls auf der Äquivalenz von räumlicher Nähe und semantischer Ähnlichkeit beruht, ist die Abfrage, wie ähnlich sich zwei Suchbegriffe sind. Als Beispiel sollen hier zwei Abfragen gegenübergestellt werden, einerseits die Ähnlichkeit von „poésie“ und „littérature“, andererseits diejenige von „prose“ und „littérature“.

Abfrage: „[prose_nom', 'littérature_nom']“

Ergebnis: 0.511518681366

Abfrage: „[poésie_nom', 'littérature_nom']“

Ergebnis: 0.714615326722

Hier zeigt das Ergebnis, dass sich offenbar Dichtung und Literatur semantisch ähnlicher sind als Prosa und Literatur – was einleuchtet, wenn man bedenkt, dass nur wenig Dichtung als nicht-literarisch, dagegen sehr viel Prosa als nicht-literarisch gilt.

3.2 Umgang mit Mehrdeutigkeit: Disambiguierung

Schon bei der Ähnlichkeitssuche wird deutlich, dass der wesentliche Gedanke bei Abfragen auf *WEM* ist, dass sich semantische Relationen in mathematische Basisoperationen übersetzen lassen. Ein weiteres eindrückliches Beispiel hierfür ist auch das Phänomen der Disambiguierung von mehrdeutigen Wörtern.²⁹

Zur Veranschaulichung soll hier das Wort „morceau“ dienen, das im Französischen sowohl ein Musikstück als auch ein Fragment eines Objektes bezeichnet. Nutzt man eine einfache Abfrage nach den Worten, die „morceau“ ähnlich sind, erhält man im Fall des *Wikipedia*-Modells eine Wortliste, die ganz auf die musikalische Bedeutung fokussiert ist.

Abfrage: „[positive = 'morceau_nom']“

Ergebnis: {"chanson_nom": 0.651456, "intro_nom": 0.612073, "sample_nom": 0.605997, "album_nom": 0.590941, "remix_nom": 0.588218}

Formuliert man allerdings „ $V(\text{morceau}) - V(\text{chanson})$ “, subtrahiert also gewissermaßen vom Vektor für „morceau“ den Vektor für das ähnlichste Wort mit musikalischem Wortsinn, erhält man Wortlisten, die nur Wörter beinhalten, die dem anderen, nicht-musikalischen Wortsinn entsprechen:

Abfrage: „[positive = 'morceau_nom', negative = 'chanson_nom']“

Ergebnis: {"lamelle_nom": 0.389208, "fragment_nom": 0.379389, "découpe_nom": 0.377873, "copeau_nom": 0.349306, "clou_nom": 0.330705, "débris_nom": 0.330573}

Die kontextabhängige Modellierung von Wortbedeutung in *WEM* ist ebenfalls Gegenstand der Forschung.³⁰

3.3 Analogiebeziehungen

Über die einfachen Ähnlichkeitsbeziehungen und die Disambiguierung durch Subtraktion hinaus konnte gezeigt werden, dass *WEM* auch in erstaunlich präziser Weise Analogiebeziehungen kodieren. Damit ist gemeint, dass in einem *WEM* nicht nur die räumliche Nähe zwischen Wörtern informationstragend ist, sondern auch die relative Position von Wörtern zueinander relevante Informationen kodiert, nämlich Bedeutungsrelationen. An einem Beispiel aufgezeigt: Richtung und Länge der Linie, die von „pommier“ („Apfelbaum“) zu „pomme“ führt, sind

²⁹Hinrich Schütze, „Automatic word sense discrimination“, in: *Computational Linguistics* 24/1 (1998), 97–124; zum konzeptuellen und mathematischen Hintergrund siehe auch Widdows (Anm. 4), 208–216.

³⁰Bspw. Stefan Thater/Hagen Fürstenau/Manfred Pinkal, „Word Meaning in Context. A Simple and Effective Vector Model“, in: *Proceedings of the 5th International Joint Conference on Natural Language Processing*, Chian Mai, Thailand, 2011, 1134–1143.

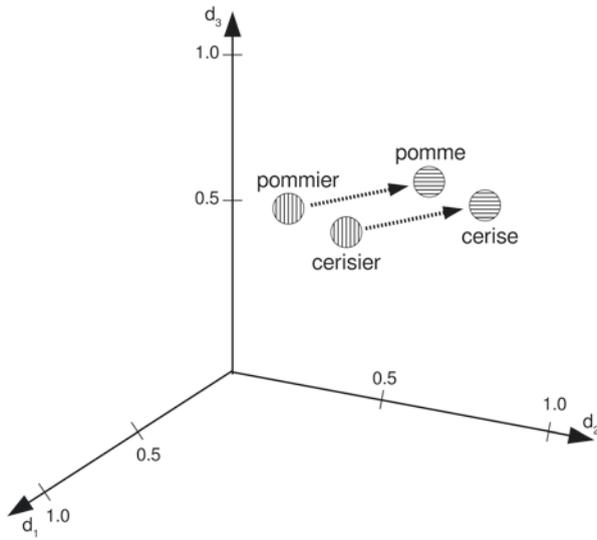


Abb. 3 Illustration der Analogiebeziehungen im Vektorraum des WEM³¹

nahezu identisch zu Richtung und Länge der Linie, die von „poirier“ (‚Birnbäum‘) zu „poire“ (‚Birne‘) führt oder von „cerisier“ (‚Kirschbaum‘) zu „cerise“ (‚Kirsche‘). Das bedeutet, dass diese Linie die Relation „Baum / Frucht“ kodiert. Abb. 3 illustriert das Prinzip dieser Analogiebeziehung.

In gleicher Weise funktioniert diese Analogiebeziehung auch für derivationale Phänomene, beispielsweise sind Richtung und Länge der Linie, die von „gehen“ zu „er ging“ führt, nahezu identisch für alle anderen Grundformen von Verben und ihre Ableitung in der dritten Person Singular der einfachen Vergangenheit. Dabei ist ein solcher Vektor nicht mit einer der Dimensionen im WEM identisch (das würde voraussetzen, dass jede Dimension von vorneherein sinnttragend ist und es nur eine fest begrenzte Zahl von semantischen Relationen geben kann), sondern seine Richtung ist als eine gewichtete Kombination aller Dimensionen des Vektorraums zu verstehen.

Auf der mathematischen Ebene entspricht der Analogierelation eine Kombination von Addition und Subtraktion von Vektoren, beispielsweise: „ $V(\text{pommier}) + V(\text{cerise}) - V(\text{pomme}) = V(\text{cerisier})$ “. Formuliert man im französischen Wikipedia-Modell die entsprechende Abfrage, erhält man in der Tat das der Analogie entsprechende Ergebnis:

Abfrage: „[positive = ['pommier_nom', 'cerise_nom'], negative = ['pomme_nom']]“
 Ergebnis: {"cerisier_nom": 0.545, "prunier_nom": 0.514, "aubépine_nom": 0.498, "prunus_nom": 0.496, "sorbier_nom": 0.494}

³¹ Bildlizenz: *Creative Commons Attribution International 4.0 (CC-BY)*.

Hier erhält man also das richtige Ergebnis: „cerisier“ (‚Kirschbaum‘) ist mit knappem Abstand das am nächsten liegende Wort. In der gleichen Weise funktionieren auch die in diesem Kontext häufig erwähnten Abfragen, die sich auf die Hauptstädte verschiedener Länder beziehen.

So einfach die bisher illustrierten Abfragemöglichkeiten erscheinen mögen, so mächtig sind sie, wenn man sie in übergeordneten Kontexten nutzt, um sprachliche Kompetenz zu modellieren und auf diese Weise in automatische Verarbeitungsroutinen zu integrieren, die dadurch nicht mehr darauf beschränkt sind, nur auf der Textoberfläche zu operieren, obwohl sie ohne unmittelbaren menschlichen Input ablaufen. Dies wird sich in den nächsten beiden Abschnitten zeigen.

4 WEM als Komponenten in übergeordneten Analyseverfahren

Im Folgenden werden zwei Anwendungsfelder für *WEM* dargestellt, in denen diese Modelle gewissermaßen als Hilfstechnologie für andere, quantitative Methoden der Analyse literarischer Texte dienen: Erstens kann man die Messung der Ähnlichkeit von Wörtern nutzen, um die Kohärenz von Topics zu messen, wie sie beim einleitend bereits erwähnten *Topic Modeling* entstehen. Zweitens kann man ein *WEM* dazu nutzen, um ein Sentiment-Lexikon für die Sentiment-Analyse zu erstellen (oder ein vorhandenes Lexikon zu erweitern oder zu modifizieren), indem man die Achse der negativen oder positiven Polarität im *WEM* ermittelt. (Im nachfolgenden Abschnitt werden dann Verfahren erläutert, mit denen direkt genuine Fragestellungen der Digitalen Literaturwissenschaften bearbeitet werden können.)

4.1 WEM als Grundlage für Maße der Topic-Kohärenz

Topic Modeling ist eine unüberwachte Methode der quantitativen Textanalyse, die dafür verwendet wird, latente semantische Strukturen in umfangreichen Textsammlungen zu entdecken, ohne dass dabei lexikalische Ressourcen wie beispielsweise elektronische Wörterbücher notwendig sind.³² Auf einer technischen Ebene ist ein ‚Topic‘ als Wahrscheinlichkeitsverteilung von Wörtern und ein ‚Dokument‘ als Wahrscheinlichkeitsverteilung von Topics repräsentiert. In der Praxis werden einzelne Wörter mit den höchsten Wahrscheinlichkeitswerten in einem Topic als in der einen oder anderen Weise semantisch verbunden verstanden, wobei die gemeinsame semantische Basis eines Topics bei der Modellierung von literarischen Texten ein abstraktes Thema, aber auch ein Setting, ein erzählerisches Motiv, eine soziale Gruppe oder ein Set rhetorischer oder argumentativer Verfahren sein kann.

³² Siehe Blei u. a. sowie Blei (jeweils Anm. 2).

Dies bedeutet, dass *Topic Modeling* selbst ein wichtiger Teil des Methodenrepertoires der quantitativen Semantik ist, weil es den Zugang zu den allgemeinen Themen und Motiven einer Textsammlung erlaubt. Allerdings haben *WEM* auch noch eine spezifischere Rolle in diesem Zusammenhang zu spielen. In der Tat liegt eine der Herausforderungen des *Topic Modelings* darin, Wege zur Evaluation von *Topic Models* zu finden, die mit bestimmten Parametern und bestimmten Textdaten erstellt wurden. Ein wichtiger Teilaspekt der Modellqualität von *Topic Models* ist dabei die Topic-Kohärenz.³³ Und *WEM* können für die Evaluation der Topic-Kohärenz eingesetzt werden. Der Vorteil gegenüber intrinsischen Maßen liegt darin, dass hier wirklich neue Information für die Evaluation genutzt wird; und der Vorteil gegenüber manch anderen externen Verfahren liegt darin, dass für eine bestimmte Sprache, Epoche und/oder Gattung ein angemessenes Messverfahren eingesetzt werden kann, ohne dass spezielle Sprachressourcen dafür notwendig sind, allerdings unter der (nicht unproblematischen) Voraussetzung, dass für die Epoche und Gattung ausreichende Textmengen für das Training eines *WEM* zur Verfügung stehen.

Das Verfahren beruht auf dem schlichten Prinzip, dass für jeden Topic die wichtigsten n Wörter ausgewählt werden und die durchschnittliche Nähe der Wörter zueinander mit einem *WEM* berechnet wird. Daraus ergibt sich dann für jeden Topic ein Kohärenz-Score. Durchschnitt und Verteilung dieser einzelnen Kohärenz-Scores kann man als Indikatoren für die Kohärenz des Topic-Modells insgesamt interpretieren.

Rang	Score	Topic-Wörter
1	0,519	Père mère enfant fille fils sœur frère maman maison famille oncle parents
2	0,435	Roi duc reine prince princesse palais royal garde empereur souverain seigneur fils
19	0,290	Table manger verre boire vin cuisine bon servir bouteille eau assiette repas
20	0,289	Porte pièce maison fenêtre mur escalier ouvrir chambre entrer étage couloir main
64	0,218	Dragon démon sorcier mage magie fois mort monde nouveau pouvoir talari passer
65	0,217	Rue maison passer porte ville regarder place café gens odeur manger sentir
97	0,151	Citoyen main ami voix porte soldat entendre ordre mouron rue air nom
98	0,150	Jeune falloir devenir manière statue dan géant nath finir peur odeur ville

³³Für einen Überblick siehe Michael Röder/Andreas Both/Alexander Hinneburg, „Exploring the Space of Topic Coherence Measures“, in: *Proceedings of the eight International Conference on Web Search and Data Mining, Shanghai, February 2–6 (2015)*, http://svn.aksw.org/papers/2015/WSDM_Topic_Evaluation/public.pdf (letzter Aufruf 8.5.2018); David Newman/Jey Han Lau/Karl Grieser u. a., „Automatic Evaluation of Topic Coherence“, in: *Human Language Technologies. The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT'10*, Stroudsburg, PA, 2011, 100–108, <http://dl.acm.org/citation.cfm?id=1857999.1858011>.

Die vorstehende Tabelle zeigt einen Ausschnitt der Ergebnisse, die auf der Grundlage des französischen *Wikipedia*-Modells für ein Topic Model des französischen Romans des 20. Jahrhunderts berechnet wurden. Gezeigt werden Rang und Wert des Topics nach seiner *WEM*-basierten Kohärenz und die 12 wichtigsten Wörter des Topics, die in die Bewertung eingeflossen sind. Eine formale Evaluation dieses Verfahrens steht noch aus, die bisherigen Ergebnisse können aber zumindest als vielversprechend gelten.

4.2 *WEM* für die Sentiment-Analyse

Die Sentiment-Analyse ist eine Methode der Textanalyse, mit der ein besonders schwer fassbarer Aspekt der Bedeutung von Sätzen oder Texten erhoben werden kann, nämlich die mehr oder weniger positive oder negative Einstellung, die in ihnen jeweils zum Ausdruck kommt.³⁴ Diese Art der Textanalyse wird vielfältig genutzt, um große Mengen von Produkt-Rezensionen oder Einträge in den sozialen Medien zu untersuchen. Die Sentiment-Analyse wird jedoch auch für die Analyse literarischer Texte verwendet, insbesondere von Erzähltexten.³⁵ Um eine solche Analyse durchführen zu können, ist ein sogenanntes Sentiment-Lexikon nützlich bzw. für diktionsbasierte Verfahren unerlässlich. Ein solches Lexikon beinhaltet (im einfachsten Fall) Einträge für einzelne Wörter und assoziiert jedes Wort mit einem bestimmten positiven oder negativen Wert, der ausdrückt, wie ausgeprägt positiv oder negativ ein Begriff ist. Beispielsweise würde das Wort „schön“ einen stark positiven Wert zugeordnet bekommen, das Wort „schrecklich“ einen deutlich negativen Wert, und ein Wort wie „blau“ vielleicht einen neutralen oder leicht positiven Wert.

Die Herausforderung, wenn man diese Art von Analysen auf historische Texte anwenden möchte, liegt darin, dass es für frühere Epochen in der Regel keine angemessenen Sentiment-Lexika gibt. ‚Angemessen‘ bedeutet in diesem Zusammenhang, dass die Polaritäts-Werte nicht von der Verwendung oder dem Verständnis der Wörter im 21. Jahrhundert abgeleitet sind, sondern korrekt wiedergeben, wie das Wort in der jeweiligen Epoche von den Lesern wahrgenommen wurde. Gewichtige Hindernisse hierfür sind nicht nur, dass die manuelle Annotation von Wörtern mit Polaritätswerten aufwendige experimentelle Setups erfordert, sondern vor allem auch, dass für historische Epochen keine authentischen Sprecher befragt werden können.

Mit einem *WEM* des Vokabulars einer historischen Textsammlung ist es aber möglich, automatisch positive oder negative Werte für Wörter aus ihrer Position im Vektorraum abzuleiten. Man benötigt hierfür lediglich eine kleine Anzahl an unstrittig deutlich positiven und deutlich negativen Ausgangswörtern (in der Wahl dieser Ausgangswörter liegt zugegebenermaßen auch die methodologische Crux

³⁴Für einen Überblick, siehe Pang/Lee (Anm. 3).

³⁵Siehe bspw. Mohammad oder Jannidis u. a. (jeweils Anm. 3).

und potenzielle Schwäche des Verfahrens). Mithilfe dieser Ausgangswörter lässt sich eine Achse durch den vieldimensionalen Vektorraum spannen, die der Polarität entspricht. Die Position der Wörter im Vokabular relativ zu dieser Achse lässt sich dann messen und dafür nutzen, um einem Wort einen mehr oder weniger ausgeprägten positiven oder negativen Wert zuzuordnen.

Bei der Beispielanalyse waren positive Ausgangswörter hier „excellent“, „remarquable“, „exceptionnel“, „étonnant“, „extraordinaire“ und „admirable“; negative Ausgangswörter waren „horrible“, „épouvantable“, „effroyable“ und „affreux“. Bemerkenswert sind hingegen die Wörter auf den jeweils nachfolgenden Rängen, die in der folgenden Tabelle dargestellt sind und die ganz zweifelsfrei deutlich positive bzw. negative Wörter sind. Ohne nennenswerten manuellen Annotationsaufwand wurden in diesem Fall 1000 Adjektive automatisch mit solchen Polaritätswerten ausgezeichnet. Die folgende Tabelle zeigt einen Ausschnitt der Ergebnisse (die jeweils 10 Wörter mit den extremsten Werten):

Polarität	Wort	Polarität	Wort
0,467	remarquable_adj	-0,385	sinistre_adj
0,407	admirable_adj	-0,388	terrible_adj
0,405	exceptionnel_adj	-0,393	ignoble_adj
0,380	incomparable_adj	-0,409	hideux_adj
0,370	étonnant_adj	-0,443	abominable_adj
0,356	merveilleux_adj	-0,466	effroyable_adj
0,346	superbe_adj	-0,490	épouvantable_
0,334	excellent_adj	-0,503	adj
0,307	magnifique_adj	-0,506	affreux_adj
0,307	charmant_adj	-0,515	horrible_adj
			atroce_adj

Mithilfe dieser Art von Polaritätstabellen lässt sich nun jeder beliebige Satz durchgehen und für jedes Inhaltswort, das in der Tabelle vorhanden ist, der Sentiment-Wert nachschlagen. In einem zweiten Schritt kann man dann Durchschnittswerte für jeden Satz oder Absatz bilden und feststellen, ob es sich um eine insgesamt eher positive oder negative Passage handelt. Die folgenden drei (erfundenen) Sätze sind mit authentischen, mit dem *Wikipedia*-Modell erstellten Sentiment-Tabellen annotiert worden:

- „L’homme poussa un cri épouvantable avant de mourir.“ (,Der Mann stieß einen schrecklichen Schrei aus, bevor er starb.) => -0,17
- „La gracieuse virtuosité des artistes faisait son bonheur.“ (,Die grazile Virtuosität der Künstler machte ihn glücklich.) => +0,18
- „Dans la salle était une vieille chaise en bois vert.“ (,In dem Raum befand sich ein alter Stuhl aus grünem Holz.) => 0,02

Neben der hier gezeigten Generierung eines völlig neuen Sentiment-Lexikons hat das Verfahren vor allem in zwei davon leicht abweichenden Szenarien Potenzial: Erstens lassen sich auf diese Weise vorhandene, häufig nicht sehr umfangreiche

Sentiment-Lexika datengetrieben erweitern, d. h. um weitere Wörter mit Polaritätswerten ergänzen. Hier ist durch den möglichen Abgleich mit vorhandenen Polaritätswerten für einen substanziellen Teil des Vokabulars mit einer besonders hohen Güte der Ergebnisse zu rechnen. Zweitens lassen sich vorhandene Lexika auf historischer Datengrundlage auf eine andere Epoche anpassen, indem die vorhandenen Werte durch datengetrieben ermittelte Werte modifiziert werden.³⁶

5 Eigenständige Methoden auf Grundlage von WEM

Zuletzt sollen nun noch ebenfalls unter Nutzung basaler Abfragemöglichkeiten solcher WEM, aber über die im letzten Abschnitt beschriebenen Hilfsfunktionen hinausgehende, stärker autonome Anwendungsgebiete thematisiert werden. Hier geht es nun auch darum Fragen zu bearbeiten, die bis in die jüngste Zeit der ureigenen Domäne der literaturwissenschaftlichen Interpretation zugeordnet worden sind. Protagonisten solcher Arbeiten, die WEM aktuell für geisteswissenschaftliche und insbesondere ideengeschichtliche und literaturhistorische Fragestellungen einsetzen, sind Ben Schmid und Ryan Heuser.³⁷

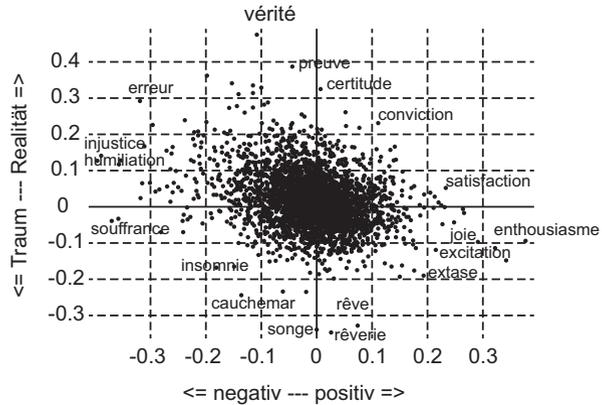
5.1 Semantische Dimensionen (Generalisierung der Polaritätsanalyse)

Das Prinzip, nach dem die Polarität von Begriffen für die Sentiment-Analyse funktioniert, kann man jenseits der spezifischen Dichotomie positiv/negativ generalisieren. Auf diese Weise lässt sich ein weiterer Anwendungsbereich der WEM für die Literatur- und Ideengeschichte erschließen. Hier wird die konzeptuelle Achse nicht nur im Sinne der positiven oder negativen Polarität durch den Vektorraum gezogen, sondern auch im Sinne bestimmter semantischer Oppositionen. Ryan Heuser spricht hier von „axes of meaning“ und nutzt sie für

³⁶Siehe Paul Cook/Suzanne Stevenson, „Automatically Identifying Changes in the Semantic Orientation of Words“, in: *Proceedings of the International Conference on Language Resources and Evaluation*, Valetta, Malta, 2010, 28–34, <http://www.cs.toronto.edu/~pcook/CookStevenson2010.pdf> (letzter Aufruf 8.5.2018).

³⁷Überblicksbeiträge sind: Ben Schmid, „Vector Space Models for the Digital Humanities“, in: *Ben's Bookworm Blog* (2015), <http://bookworm.benschmidt.org/posts/2015-10-25-Word-Embeddings.html>; Ryan Heuser, „Word Vectors in the Eighteenth Century“, in: *Adventures of the Virtual* (2016), <http://ryanheuser.org/word-vectors/> (letzter Aufruf 8.5.2018); Ders., „Word Vectors in the Eighteenth Century“, in: *Digital Humanities 2017: Conference Abstracts*, Montréal 2017, 256–260. Siehe auch: Siobhán Grayson/Maria Mulvany/Karen Wade u. a., „Exploring the Role of Gender in 19th Century Fiction Through the Lens of Word Embeddings“, in: *LDK 2017: Language, Data and Knowledge, Galway, Ireland, 19–20 June 2017*, Cham, CH, 2017, 358–364, https://doi.org/10.1007/978-3-319-59888-8_30.

Abb. 4 Wörter in den Dimensionen negativ/positiv (horizontale Achse) und Traum/Realität (vertikale Achse) im Roman-Modell³⁹



die ideengeschichtliche Analyse des Schrifttums des englischen achtzehnten Jahrhunderts, indem er historisch abgesicherte Dichotomien wie „abstract vs. concrete“ oder „simplicity vs. refinement“ unter Verwendung von *WEM* analysiert.³⁸ In einem weiteren Schritt lassen sich nach diesem Prinzip darüber hinaus nicht nur eindimensionale, sondern auch zweidimensionale semantische Räume definieren, also zwei solche semantische Achsen ziehen. Dies erlaubt es dann, Wörter in einem zweidimensionalen Raum zu lokalisieren und Zusammenhänge zwischen den beiden semantischen Dimensionen zu explorieren, wie sie sich in einer bestimmten Textsammlung zeigen.

Abb. 4 zeigt eine solche Visualisierung unter Nutzung des Roman-Modells, wobei die horizontale Achse der positiven/negativen Polarität entspricht, die vertikale Achse aber der Dichotomie Traum/Realität. Zwar zeigt sich hier keine deutliche Korrelation zwischen den beiden Achsen, wie sie Ryan Heuser beispielsweise im 18. Jahrhundert für „virtue“ und „simplicity“ festgestellt hat. Von besonderem Interesse sind dort aber auch diejenigen Wörter, die dieser Korrelation widerstreben: So gilt „science“ dem englischen achtzehnten Jahrhundert offenbar als eine raffinierte Tugend, „murder“ hingegen als ein eher einfaches Laster. Die Darstellung zum französischen Roman des 20. Jahrhunderts erlaubt ebenfalls einen Blick in die Zusammenhänge der beiden semantischen Dimensionen; bei ähnlicher Polarität unterscheiden sich so bestimmte Wörter in Bezug auf das Ausmaß ihrer Assoziation mit Traum oder Realität: so für positive Polarität „extase“ (,Extase‘) und „foi“ (,Glaube‘) oder für negative Polarität „souffrance“ (,Leiden‘) und „malheur“ (,Unglück‘). Nicht vergessen werden sollte bei der Betrachtung solcher Darstellungen, dass sie nicht zeitgenössische Sprachkompetenz modellieren, sondern die konzeptuelle Struktur einer spezifischen literarischen Gattung, hier des Romans des 20. Jahrhunderts. Die gleiche Analyse, berechnet für das *Wikipedia*-Modell oder für Texte aus der Aufklärung, würde deutlich andere Ergebnisse zeigen.

³⁸ Heuser 2016 (Anm. 37).

³⁹ Bildlizenz: *Creative Commons Attribution International 4.0 (CC-BY)*.

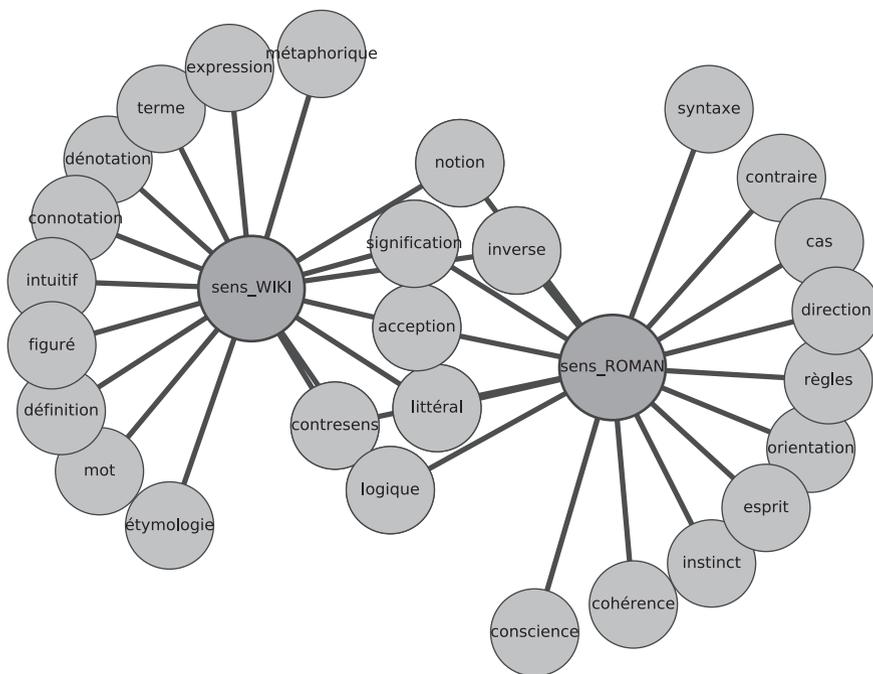


Abb. 5 Vergleichendes konzeptuelles Netzwerk für das Wort „sens“ („Bedeutung“) in den auf der Wikipedia und der Romansammlung beruhenden Modellen⁴⁰

5.2 Konzeptuelle Strukturen literarischer Gattungen

Eine solche vergleichende Perspektive nimmt die folgende Analyse ein. Hier geht es darum, die latente, konzeptuelle Struktur einer literarischen Domäne, beispielsweise einer bestimmten Gattung in einer bestimmten Epoche, vergleichend zu explorieren. Im vorliegenden Setup werden hierfür zwei *WEM* miteinander verglichen: einerseits das *Wikipedia*-Modell (als Vertreter aktueller sachbezogener Textsorten), andererseits das Roman-Modell (als Vertreter einer spezifischen literarischen Gattung, allerdings hier ohne präzise chronologische Eingrenzung). Das Roman-Modell nicht nur für sich stehend zu explorieren (unter anderem mit den im Abschn. 3 dargelegten Abfrageverfahren), sondern es hier mit einem weiteren Modell zu kontrastieren, macht die latente konzeptuelle Struktur der Gattung deutlich sichtbar.

Abb. 5 zeigt die 15 Wörter, die dem Wort „sens“ („Sinn“, „Bedeutung“) im Roman-Modell und im *Wikipedia*-Modell am ähnlichsten sind. Die Worte in der Mitte (rot) teilen beide Modelle, die außen liegenden Worte (türkis) hingegen sind jeweils nur im einen oder anderen Modell unter den ähnlichsten 15 Wörtern ver-

⁴⁰ Bildlizenz: *Creative Commons Attribution International 4.0 (CC-BY)*.

treten. Was hier deutlich wird, ist das sehr systematische konzeptuelle System der *Wikipedia*, das mit „sens“ wichtige Konzepte wie ‚Definition‘, ‚Etymologie‘ und ‚Denotation‘ verbindet. Im Gegensatz dazu stehen im Romanmodell Begriffe im Vordergrund, die „Sinn“ mit menschlicher Kognition und menschlichem Handeln verbinden, so ‚Bewusstsein‘, ‚Instinkt‘ und ‚Geist‘.

5.3 Die Entwicklung von Wortbedeutungen nachvollziehen

Ein weiterer Anwendungsbereich von *Word Embedding Models* liegt im Bereich der historischen Semantik und der Ideengeschichte. Hier haben Hamilton u. a. eine Methodik vorgeschlagen, mit der semantischer Wandel daran ablesbar wird, wie sich ein Begriff nach und nach von verwandten Begriffen entfernt, um sich anderen Begriffen anzunähern.⁴¹ Dies kann man im Sinne der distributionellen Semantik als Hinweis auf die semantische Dissoziation des Begriffs von den Bedeutungskomponenten der ersten Gruppe und als eine Assoziation mit denjenigen der zweiten Gruppe verstehen.

Auch Ryan Heuser nutzt diese Methodik, um eine Reihe von Hypothesen zur Ideengeschichte des 18. Jahrhunderts von Reinhart Koselleck und Raymond Williams zu überprüfen.⁴² Er kann beispielsweise zeigen, dass das Wort „industry“ zwischen 1700 und 1800 seinen Abstand zu bestimmten Worten vergrößert, darunter „skill“, „dexterity“ und „prudence“. Gleichzeitig verringert sich der Abstand von „industry“ zu bestimmten anderen Worten, darunter „agriculture“, „manufacture“, „idleness“, „sobriety“. In anderen Worten: Es bestätigt sich hier datenbasiert, dass die Bedeutung von „industry“ sich von der persönlichen Kompetenz und Betriebsamkeit weg und zur mechanisierten Großproduktion hin entwickelt.

6 Fazit

In den vorangegangenen Abschnitten sind einfache Abfragemöglichkeiten von *WEM*, der Einsatz von *WEM* als unterstützende Technologien in bestimmten Analyseverfahren und die unmittelbare Nutzung von *WEM* für Fragestellungen aus den digitalen Literaturwissenschaften thematisiert worden. Methodisch spannend daran ist vor allem, wie weitreichende Anwendungen auch auf der Grundlage relativ basal erscheinender Abfragen möglich werden, wenn man nur eine geeignete Operationalisierung der jeweils infrage stehenden Problemstellung findet. Nun ist es jedoch an der Zeit, zur eingangs formulierten Fragestellung zurückzukehren und

⁴¹William L. Hamilton/Jure Leskovec/Dan Jurafsky, „Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change“, in: *arXiv:1605.09096 [cs.CL]* (30. Mai 2016), <http://arxiv.org/abs/1605.09096> (letzter Aufruf 8.5.2018).

⁴²Heuser 2017 (Anm. 37).

eine kritische Einschätzung dazu zu formulieren, ob mit den *WEM* und weiteren aktuellen semantischen Technologien die etablierte hermeneutisch vorgehende Literaturinterpretation tatsächlich ernsthaft herausgefordert ist.

Trotz des meines Erachtens erheblichen und nachweislichen Potenzials dieser Methoden scheint es offensichtlich, dass dies zumindest derzeit aus mehreren Gründen noch nicht der Fall ist; Gründen allerdings, die nicht unbedingt als grundsätzliche Hindernisse zu verstehen sind, sondern die sich in den nächsten Jahren mit Sicherheit deutlich verringern werden. Zu diesen Hindernissen gehört zunächst, dass schlicht noch sehr wenig Erfahrungen mit der Anwendung von *WEM* für die digitalen Literaturwissenschaften vorliegen, nämlich – streng genommen – nur die Arbeiten von Ryan Heuser, die zudem noch nicht in ausführliche formale Publikationen gemündet sind. Dies wird sich, das hat die deutliche Präsenz des Themas auf den *Digital Humanities Conferences* 2017 in Montréal oder 2018 in Mexico City gezeigt, in den nächsten Jahren zweifellos ändern. Hier zeichnet sich ab, dass *WEM* nicht nur für sich genommen ein spannendes Betätigungsfeld für die angewandte Informatik sind, sondern auch ein wesentlicher Baustein im Methodenrepertoire der digitalen Geistes- und Literaturwissenschaften sein können, beispielsweise im Kontext von Methoden wie dem *Topic Modeling*, der Sentiment-Analyse oder für Verfahren der korpusbasierten Ideengeschichte. Wenn vorhandene Hypothesen auf breiter Datengrundlage überprüft und ggf. auch nuanciert oder infrage gestellt werden können, wie dies bereits der Fall ist, dann ist es nicht mehr weit, bis die Methoden auch helfen werden, neue verlässliche Erkenntnisse zutage zu fördern.

Ein weiteres, sicherlich weniger leicht zu behebendes Hindernis ist die derzeit mangelnde Integration von *WEM* und den Möglichkeiten, die sie mit sich bringen, mit anderen vorhandenen Technologien. *WEM* erlauben zwar mit bisher ungekannter Präzision einen Zugang zu semantischen und morpho-syntaktischen Ähnlichkeitsbeziehungen zwischen Einzelwörtern, und es lassen sich auf dieser Grundlage eine Reihe von Analyseverfahren aufbauen, die beim Umgang mit lexikalischer und textueller Bedeutung nützlich sind. Aber *WEM* sind prinzipiell relationale Systeme, die Beziehungen zwischen Wörtern, aber nicht explizit Bedeutungen von Wörtern kodieren. Zudem gilt, dass die Modellierung semantischer Kompetenz, wie sie *WEM* leisten, erst mit der Modellierung anderer sprachlicher Bereiche, wie der syntaktischen Sprachkompetenz, verbunden werden muss, damit es möglich wird, mit computergestützten Systemen den propositionalen Inhalt einzelner Sätze in sinnvoller und gewinnbringender Weise zu modellieren. Erst dann wird man an die Integration von historischem, kulturellem und ästhetischem Kontextwissen denken können, um einzelnen Sätzen oder längeren Passagen eine in einem solchen Kontext adäquate Bedeutung automatisch zuschreiben zu können. Anders ausgedrückt: Auch wenn *WEM* unbestrittenes Potenzial für die Digitalen Literaturwissenschaften besitzen, sind sie derzeit noch weit davon entfernt, die Position und Kompetenz der hermeneutischen Textinterpretation herauszufordern.

Ergänzung 2021

Welche für die *Computational Literary Studies* (CLS) besonders relevanten, neueren Entwicklungen hat es im Bereich Word Embedding Models seit 2017

gegeben? Und führen diese Entwicklungen zu einer gegenüber dem Fazit des damaligen Beitrags veränderten Einschätzung der Rolle von Word Embedding Models in den digitalen Geisteswissenschaften?

Zunächst kann man feststellen, dass die Vielfalt der verfügbaren Embeddings in den letzten fünf Jahren weiter deutlich angestiegen ist. Über die im Beitrag erwähnten, Wort-basierten Varianten von Word2Vec, wie GloVe und FastText hinaus, wurden auch die von den Embeddings repräsentierten Einheiten vielfältiger: So gibt es mittlerweile auch Embeddings für andere Einheiten, die sowohl kleiner (N-Gramme oder Buchstaben) als auch größer als das Wort (Phrasen, Sätze, Absätze oder Dokumente) sein können.

Auch die den Embeddings zugrunde liegenden Korpora sind weiter deutlich gewachsen: Während das ursprüngliche Word2Vec-Modell 2013 mit 1,6 Mrd. Wörtern trainiert wurde, werden neuere Sprachmodelle wie beispielsweise GPT-3 mit bis zu 500 Mrd. Wörtern trainiert und haben auch wesentlich mehr Dimensionen und Parameter als frühere Modelle.⁴³ Dies schlägt sich deutlich in der Performanz und Flexibilität der Modelle nieder, hat aber durchaus auch nachteilige Konsequenzen.⁴⁴

Die wichtigste konzeptuelle Neuerung ist aber wohl der neue Typus der “deep, contextualized” Embeddings. Dieser Typus ist “deep”, insofern er auf einer größeren Anzahl verdeckter Schichten in der Architektur des neuronalen Netzes beruht (während das klassische Word2Vec mit nur einer verdeckten Schicht als “shallow” beschrieben werden kann). Und dieser neue Typus ist “contextualized”, weil nicht mehr jede unterschiedliche Wortform nur einen für alle Vorkommen im Korpus geltenden, einheitlichen Vektor hat, wie dies bei Word2Vec der Fall ist. Vielmehr wird der Vektor jedes einzelnen Wortes in Abhängigkeit des jeweiligen Kontextes individuell angepasst. Dadurch können nicht nur die unterschiedlichen Bedeutung von Homographen oder von polysemen Wörtern differenziert werden, sondern es können auch vom individuellen Kontext abhängige Interpretationen beispielsweise von Pronomina oder Bedeutungsnuancen beliebiger Wörter differenziert kodiert werden.

Die ersten solchen kontextualisierten Embeddings wurden noch mit sequenziellen Trainingsmechanismen trainiert, zumeist mit Recurrent Neural Networks wie den LSTMs (Long-Short-Term-Memory), allerdings in einem bidirektionalen Modus, der es besser erlaubte, kontextabhängige Repräsentationen zu ermitteln. Eine weitere entscheidende Neuerung war dann, dies durch nicht-sequentielle Mechanismen

⁴³Tom B. Brown/Benjamin Mann/Nick Ryder u. a., „Language Models are Few-Shot Learners“, in: arXiv:2005.14165 [cs] (2020), <http://arxiv.org/abs/2005.14165>; Alec Radford/Jeff Wu/R. Child u. a., „Language Models Are Unsupervised Multitask Learners“ (2019), <https://www.semanticscholar.org/paper/Language-Models-are-Unsupervised-Multitask-Learners-Radford-Wu/9405cc0d6169988371b2755e573cc28650d14dfe> (letzter Aufruf der Links 24.10.2021).

⁴⁴Emily M. Bender/Timnit Gebru/Angelina McMillan-Major u. a., “On the Dangers of Stochastic Parrots. Can Language Models Be Too Big?”, in: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (2021), 610–23. DOI: <https://doi.org/10.1145/3442188.3445922> (letzter Aufruf 24.10.2021).

zu ersetzen, wie es von Google mit dem Transformer-basierten Modell BERT erstmals vorgeschlagen wurde.⁴⁵ Dabei kommt der sogenannte “Attention”-Mechanismus zum Einsatz, bei dem der gesamte Satzkontext auf einmal, aber pro Token mit unterschiedlichen Gewichten, für die Repräsentation der Tokens genutzt wird.⁴⁶ Das Modell wird mit zwei Tasks trainiert: dem bekannten Wortvorhersage-Task und der Aufgabe festzulegen, welcher Satz auf welchen vorigen Satz folgt. Mit diesem Verfahren können auch über größere Abstände vorhandene Abhängigkeiten zwischen Wörtern und Sätzen besser und vor allem durch Parallelisierung effizienter (und damit auch auf der Grundlage größerer Mengen an Trainingsdaten) modelliert werden. In der Summe führen diese Neuerungen zu wesentlichen Verbesserungen in der Performance dieser kontextualisierten Embeddings in einer großen Bandbreite von Tasks: sowohl für relative generische Standard-Tasks (wie POS-Tagging oder Named Entity Recognition) als auch für komplexere Aufgaben (wie *Question Answering* oder maschinelle Übersetzung).⁴⁷

Sind diese neuen, kontextualisierten Embeddings aber auch tatsächlich für die CLS relevant und leisten sie hier mehr oder Anderes als die klassischen Word Embeddings? Ted Underwood ist mit guten Gründen skeptisch, wenn es um eine Aufgabe wie die Klassifikation umfangreicher literarischer Texte nach ihrem Genre geht.⁴⁸ Andere Anwendungen, die von der Information über Beziehungen zwischen Wörtern in Texteinheiten von der Größenordnung eines Absatzes profitieren, scheinen besser geeignet. In der Tat gibt es entsprechende Anwendungen auch bereits in den CLS: Beispiele sind die Identifikation von Ereignissen in fiktionaler Prosa oder die Erkennung direkter Rede in mehrsprachigen literarischen Romankorpora.⁴⁹ Die Interpretation literarischer Werke werden auch die kontextualisierten Word Embeddings nicht übernehmen; aber sie werden bei vielfältigen Fragestellungen in den CLS einen wichtigen Beitrag dazu leisten, unseren Zugriff auf umfangreiche Textbestände zunehmend nuanciert zu gestalten.

⁴⁵ Jacob Devlin/Ming-Wei Chang/Kenton Lee u. a., „BERT. Pre-training of Deep Bidirectional Transformers for Language Understanding“, in: arXiv:1810.04805 [cs] (2018), <http://arxiv.org/abs/1810.04805> (letzter Aufruf 24.10.2021).

⁴⁶ Ashish Vaswani/Noam Shazeer/Niki Parmar u. a., “Attention is all you need“, in: *Proceedings of the 31st International Conference on Neural Information Processing Systems* (2017), 6000–6010.

⁴⁷ Eine anspruchsvolle, aber kleinschrittige und informativ illustrierte Erklärung des Transformer-Modells ist: Jay Alammar, “*The Illustrated Transformer*“, in: *Jalammar Blog* (2018), <http://jalammar.github.io/illustrated-transformer/> (letzter Aufruf 24.10.2021).

⁴⁸ Ted Underwood, “Do Humanists Need BERT?“, in: *The Stone and the Shell* (2019), <https://tedunderwood.com/2019/07/15/do-humanists-need-bert/> (letzter Aufruf 24.10.2021).

⁴⁹ Matthew Sims/Jong Ho Park/David Bamman, “Literary Event Detection“, in: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (2019), 3623–34. <https://doi.org/10.18653/v1/P19-1353>; Joanna Byszuk/Michal Woźniak/Mike Kestemont u. a., “Detecting Direct Speech in Multilingual Collection of 19th-Century Novels“, in: *Proceedings of LT4HALA 2020 – 1st Workshop on Language Technologies for Historical and Ancient Languages* (2020), 100–104, <https://www.aclweb.org/anthology/2020.lt4hala-1.15>.

Eine letzte Entwicklung sei noch genannt: So ist der Einsatz von BERT und Deep Learning allgemein seit 2017 auch für digitale Geisteswissenschaftler:innen einfacher geworden, beispielsweise durch die Entwicklung von *Keras*, einem Python-Interface für das Machine-Learning-Framework *Tensorflow*.⁵⁰

Anhang: Verwendete Tools, Code und Daten

Die hier vorgestellten Analysen wurden unter Verwendung der *gensim*-Bibliothek für Python durchgeführt, die von Radim Rehurek entwickelt wird.⁵¹ Mit dieser Bibliothek wurden zwei *Word Embedding Models* erstellt:

- Das erste Modell basiert auf einer Sammlung französischer Romane des 20. Jahrhunderts und deckt die Zeit von 1900 bis 2010 ab, wobei eine große Bandbreite von Autorinnen und Autoren und Romangattungen vertreten ist. Die Sammlung enthält ungefähr 1200 Romane und ca. 60 Mio. Tokens. Das Modell wurde mit der *SGNS*-Architektur (*Skip-Gram with Negative Sampling*) erstellt. Weitere Parameter waren ein Modell mit 300 Dimensionen, ein Kontextfenster von 6+6 Wörtern und eine minimale Worthäufigkeit von 50.
- Das zweite Modell wurde von der französischsprachigen *Wikipedia* (XML-Kopie vom 12. Juni 2017) abgeleitet. Dieser Datensatz enthält 1881 Mio. Artikel und ungefähr 764 Mio. Tokens. Das Modell wurde ebenfalls mit der *SGNS*-Architektur erstellt. Weitere Parameter waren ein Modell mit 300 Dimensionen, ein Kontextfenster von 6+6 Wörtern und eine minimale Worthäufigkeit von 100.

Der verwendete Code ebenso wie eine Auswahl an abgeleiteten Daten und einige Visualisierungen sind im Projekt-Repository der *CLiGS*-Gruppe abgelegt.⁵²

Literatur⁵³

Alammar, Jay, "The Illustrated Transformer", in: *Jalammar Blog* (2018), <http://jalammar.github.io/illustrated-transformer/> (letzter Aufruf 24.10.2021).

⁵⁰ François Chollet u. a., "Keras". Version 2.4 (2020), <https://keras.io> (letzter Aufruf der Links 24.10.2021).

⁵¹ Siehe <https://radimrehurek.com/gensim/> sowie Radim Rehurek/Petr Sojka, „Software Framework for Topic Modelling with Large Corpora“, in: *The LREC 2010 Workshop on New Challenges for NLP Frameworks*, Malta 2010, 45–50, <http://is.muni.cz/publication/884893/enhurek> (jeweils letzter Aufruf 6.5.2018).

⁵² Siehe <https://github.com/cligs/projects/tree/master/2017>, <https://doi.org/10.5281/zenodo.439982>.

⁵³ Sämtliche digitalen Referenzen wurden letztmalig am 8.5.2018 eingesehen.

- Baroni, Marco/Dinu, Georgiana/Kruszewski, Germán, „Don't Count, Predict! A Systematic Comparison of Context-counting vs. Context-predicting Semantic Vectors“, in: *Proceedings of Association for Computational Linguistics (ACL)*, Bd. 1, Baltimore, MD, 2014, 238–247, <http://acl2014.org/acl2014/P14-1/pdf/P14-1023.pdf>.
- Bender, Emily M./Timnit Gebru/Angelina McMillan-Major u. a., “On the Dangers of Stochastic Parrots. Can Language Models Be Too Big?“, in: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (2021), 610–23. DOI: <https://doi.org/10.1145/3442188.3445922> (letzter Aufruf 24.10.2021).
- Blei, David M., „Probabilistic Topic Models“, in: *Communications of the ACM* 55/4 (2012), 77–84, DOI <https://doi.org/10.1145/2133806.2133826>.
- Blei, David M./Ng, Andrew Y./Jordan, Michael I., „Latent Dirichlet Allocation“, in: *Journal of Machine Learning Research* 3 (2003), 993–1022, <http://dl.acm.org/citation.cfm?id=944919.944937>.
- Blevins, Cameron, „Topic Modeling Martha Ballard's Diary“, in: *Historizing* (April 2010), <http://historizing.org/2010/04/01/topic-modeling-martha-ballards-diary/>.
- Brown, Tom B./Benjamin Mann/Nick Ryder u. a., “Language Models are Few-Shot Learners“, in: arXiv:2005.14165 [cs] (2020), <http://arxiv.org/abs/2005.14165>.
- Bullinaria, John A./Levy, Joseph P., „Extracting Semantic Representations from Word Co-Occurrence Statistics. A Computational Study“, in: *Behavior Research Methods* 39/3 (2007), 510–526, DOI <https://doi.org/10.3758/BF03193020>.
- Byszuk, Joanna/Woźniak, Michał/Kestemont, Mike u. a., “Detecting Direct Speech in a Multilingual Collection of 19th-Century Novels“, in: *Proceedings of LT4HALA 2020 – 1st Workshop on Language Technologies for Historical and Ancient Languages*. Marseille: European Language Resources Association (ELRA), 2020, 100–104, <https://www.aclweb.org/anthology/2020.lt4hala-1.15>.
- Chollet, François u. a., “Keras”. Version 2.4 (2020), <https://keras.io> (letzter Aufruf der Links 24.10.2021).
- Cook, Paul/Stevenson, Suzanne, „Automatically Identifying Changes in the Semantic Orientation of Words“, in: *Proceedings of the International Conference on Language Resources and Evaluation*, Valetta, Malta, 2010, 28–34, <http://www.cs.toronto.edu/~pcook/CookStevenson2010.pdf>.
- Deerwester, Scott/Dumais, Susan T./Furnas, George W. u. a., „Indexing by Latent Semantic Analysis“, in: *Journal of the American Society for Information Science* 41/6 (1990), 391–407, DOI [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<391::AID-AS11>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-AS11>3.0.CO;2-9).
- Devlin, Jacob/Ming-Wei Chang/Kenton Lee u. a., “BERT. Pre-training of Deep Bidirectional Transformers for Language Understanding“, in: arXiv:1810.04805 [cs] (2018), <http://arxiv.org/abs/1810.04805> (letzter Aufruf 24.10.2021).
- Firth, John R., „A Synopsis of Linguistic Theory, 1930–1955“, in: *Studies in Linguistic Analysis* (1957), 1–32.
- Gardenfors, Peter, *Conceptual Spaces. The Geometry of Thought*, Cambridge, MA, 1998.
- Grayson, Siobhán/Mulvany, Maria/Wade, Karen u. a., „Exploring the Role of Gender in 19th Century Fiction Through the Lens of Word Embeddings“, in: *LDK 2017: Language, Data and Knowledge, Galway, Ireland, 19–20 June 2017*, Cham, CH, 2017, 358–364, DOI https://doi.org/10.1007/978-3-319-59888-8_30.
- Hamilton, William L./Leskovec, Jure/Jurafsky, Dan, „Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change“, in: arXiv:1605.09096 [cs.CL] (30. Mai 2016), <http://arxiv.org/abs/1605.09096>.
- Harris, Zellig, „Distributional Structure“, *Word* 10/2-3 (1954), 146–62, DOI <https://doi.org/10.1080/00437956.1954.11659520>.
- Heuser, Ryan, „Word Vectors in the Eighteenth Century“, in: *Adventures of the Virtual* (2016), <http://ryanheuser.org/word-vectors/>.
- Heuser, Ryan, „Word Vectors in the Eighteenth Century“, in: *Digital Humanities 2017: Conference Abstracts*, Montréal 2017, 256–260.
- Jannidis, Fotis/Reger, Isabella/Zehle, Albin, „Analyzing Features for the Detection of Happy Endings in German Novels“, in: arXiv:1611.09028 [cs.IR] (28. November 2016), <http://arxiv.org/abs/1611.09028>.

- Joulin, Armand/Grave, Edouard/Bojanowski, Piotr u. a., „Bag of Tricks for Efficient Text Classification”, in: [arXiv:1607.01759](https://arxiv.org/abs/1607.01759) (Juli 2016), <http://arxiv.org/abs/1607.01759>.
- Landauer, Thomas K. / Dumais, Susan T., „A solution to Plato's problem. The latent semantic analysis theory of acquisition, induction, and representation of knowledge”, in: *Psychological Review* 104/2 (1997), 211–240, DOI <https://doi.org/10.1037/0033-295X.104.2.211>.
- Landauer, Thomas K./Foltz, Peter W./Laham, Darrell, „Introduction to Latent Semantic Analysis”, in: *Discourse Processes* 25 (1998), 259–284.
- Levy, Omer/Goldberg, Yoav/Dagan, Ido, „Improving Distributional Similarity with Lessons Learned from Word Embeddings“, in: *Transactions of the Association for Computational Linguistics* 3 (2015), 211–225, <https://www.transacl.org/ojs/index.php/tacl/article/view/570>.
- Maaten, Laurens van der/Hinton, Geoffrey, „Visualizing Data Using T-SNE“, in: *Journal of Machine Learning Research* 9 (2008), 2579–2605, <http://www.jmlr.org/papers/v9/vandermaaten08a.html>.
- McClelland, James L./Rumelhart, David E. (Hg.), *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*, Cambridge, MA, 1986.
- Mikolov, Tomas/Chen, Kai/Corrado, Greg u. a., „Efficient Estimation of Word Representations in Vector Space“, in: *arXiv:1301.3781 [cs.CL]* (7. September 2013), <http://arxiv.org/abs/1301.3781>.
- Mohammad, Saif, „From Once Upon a Time to Happily Ever After. Tracking Emotions in Novels and Fairy Tales“, in: *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, Stroudsburg, PA, 2011, 105–114, <http://dl.acm.org/citation.cfm?id=2107636.2107650>.
- Newman, David/Lau, Jey Han/Grieser, Karl u. a., „Automatic Evaluation of Topic Coherence“, in: *Human Language Technologies. The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT '10*, Stroudsburg, PA, 2010, 100–108, <http://dl.acm.org/citation.cfm?id=1857999.1858011>.
- Newman, David/Bonilla, Edwin V./Buntine, Wray. „Improving Topic Coherence with Regularized Topic Models,” in: *Advances in Neural Information Processing Systems*, vol. 24. Curran Associates, 2011.
- o.N., „Vector Representations of Words“, in: *Tensor Flow Tutorials* (2018), <https://www.tensorflow.org/tutorials/word2vec>.
- Pang, Bo/Lee, Lillian, „Opinion Mining and Sentiment Analysis“, in: *Found. Trends Inf. Retr.* 2/1-2 (2008), 1-135, DOI <https://doi.org/10.1561/15000000011>.
- Pennington, Jeffrey/Socher, Richard/Manning, Christopher D., *Glove. Global Vectors for Word Representation*, Stanford 2014, <http://citeseerx.ist.psu.edu/viewdoc/citations;jsessionid=B90254BA67F435112ACC1AC456222FA9?doi=10.1.1.671.1743>.
- Radford, Alec/Wu, Jeff /Child, Rewon u.a., “Language Models Are Unsupervised Multitask Learners”, Technical Report. San Francisco: OpenAI, 2019.
- Rehurek, Radim/Sojka, Petr, „Software Framework for Topic Modelling with Large Corpora“, in: *The LREC 2010 Workshop on New Challenges for NLP Frameworks*, Malta 2010, 45–50, <http://is.muni.cz/publication/884893/en>.
- Röder, Michael/Both, Andreas/Hinneburg, Alexander, „Exploring the Space of Topic Coherence Measures“, in: *Proceedings of the eight International Conference on Web Search and Data Mining, Shanghai, February 2–6* (2015), http://svn.aksw.org/papers/2015/WSDM_Topic_Evaluation/public.pdf
- Salton, Gerard/Wong, Andrew/Yang, Chung-Shu, „A Vector Space Model for Automatic Indexing”, in: *Communications of ACM* 18/11 (1975), 613–20, DOI <https://doi.org/10.1145/361219.361220>.
- Schmid, Ben, „Vector Space Models for the Digital Humanities”, in: *Ben's Bookworm Blog* (2015), <http://bookworm.benschmidt.org/posts/2015-10-25-Word-Embeddings.html>.
- Schöch, Christof, „Topic Modeling Genre. An Exploration of French Classical and Enlightenment Drama“, in: *Digital Humanities Quarterly* 11/2 (2017), <http://www.digitalhumanities.org/dhq/vol/11/2/000291/000291.html>.
- Schütze, Hinrich, „Automatic word sense discrimination“, in: *Computational Linguistics* 24/1 (1998), 97–124.

- Sims, Matthew/Jong Ho Park/David Bamman, “Literary Event Detection”, in: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (2019), 3623–34. DOI: <https://doi.org/10.18653/v1/P19-1353>.
- Steyvers, Mark/Griffiths, Tom, “Probabilistic Topic Models”, in: *Latent Semantic Analysis: A Road to Meaning*, T. Landauer, D. McNamara, S. Dennis, and W. Kintsch (Hg.). Laurence Erlbaum, 2006. <http://psiexp.ss.uci.edu/research/papers/SteyversGriffithsLSABookFormatted.pdf>.
- Stubbs, Michael, „Notes on the History of Corpus Linguistics and Empirical Semantics“, in: Marja Nenonen/Sinikka Niemi (Hg.), *Collocations and Idioms*, Joensuu, Finland, 2007, 317–329.
- Thater, Stefan/Fürstenau, Hagen/Pinkal, Manfred, „Word Meaning in Context. A Simple and Effective Vector Model“, in: *Proceedings of the 5th International Joint Conference on Natural Language Processing*, Chian Mai, Thailand, 2011, 1134–1143.
- Turney, Peter T./Pantel, Patrick, „From Frequency to Meaning. Vector Space Models of Semantics“, in: *Journal of Artificial Intelligence Research* 37 (2010), 141–188, <https://arxiv.org/abs/1003.1141>.
- Underwood, Ted, “Do Humanists Need BERT?“, in: *The Stone and the Shell* (2019), <https://tedunderwood.com/2019/07/15/do-humanists-need-bert/> (letzter Aufruf 24.10.2021).
- Vaswani, Ashish/Noam Shazeer/Niki Parmar u. a., “Attention is all you need“, in: *Proceedings of the 31st International Conference on Neural Information Processing Systems* (2017), 6000–6010.
- Widdows, Dominic, *Geometry and Meaning*, Stanford, CA, 2004.

Online-Ressourcen

- ECCO (Eighteenth-Century Collection Online), <https://quod.lib.umich.edu/e/ecco/>.
- Gensim, <https://radimrehurek.com/gensim/>.
- GitHub, CLiGS-Gruppe, <https://github.com/cligs/projects/tree/master/2017>, DOI <https://doi.org/10.5281/zenodo.439982>.
- Google-Ngram-Korpus, <http://storage.googleapis.com/books/ngrams/books/datasetsv2.html>.
- Hathi Trust, <https://www.hathitrust.org/>.
- TextGrid, Digitale Bibliothek, <https://textgrid.de/digitale-bibliothek>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Small Worlds, Beat Charts und die Netzwerkanalyse dramatischer Texte. Reflexionen aus dem *Rabbit Hole*

Peer Trilcke

1 Conceptual Issues

In *Literature, Measured* hat Franco Moretti das revolutionäre Potenzial der *Digitalen Literaturwissenschaft* in einer „radicalization of our relationship to concepts“ verortet.¹ Damit wiederholt er seine Prognose, nach der „the encounter of computation and criticism“ und also die Arbeit nach dem „measurement“-Paradigma in der Literaturwissenschaft vorerst nicht zu einem „change“ der Literaturgeschichte, sondern zu einem „change“ der „theory of literature“ führen werde:

I assumed, like so many others, that the new approach would change the history, rather than the theory of literature; and, ultimately, that may still be the case. But as the logic of research has brought us face to face with conceptual issues, they should openly become the task of the day countering the pervasive clichés on the simple-minded positivism of digital humanities. Computation has theoretical consequences – possibly, more than any other field of literary study. The time has come, to make them explicit.²

¹Franco Moretti, „Literature, Measured“, in: *Literary Lab Pamphlet* 12 (April 2016), <https://litlab.stanford.edu/LiteraryLabPamphlet12.pdf> (letzter Aufruf 30.5.2019), 5.

²Franco Moretti, „Operationalizing“. Or, the function of measurement in modern literary theory“, in: *Literary Lab Pamphlet* 6 (Dezember 2013), 9, <https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf> (letzter Aufruf 30.5.2019).

Der Autor steht hier stellvertretend für die *dlina*-Arbeitsgruppe. An den hier präsentierten und diskutierten Analysen waren insbesondere Frank Fischer, Christopher Kittel, Mathias Göbel und Dario Kampkaspar beteiligt.

P. Trilcke (✉)
Universität Potsdam, Potsdam, Deutschland
E-Mail: trilcke@uni-potsdam.de

Was im Folgenden, einige Studien der *dlina*-Arbeitsgruppe³ rekapitulierend, dargestellt wird, ist zugleich bekräftigender wie problematisierender Kommentar zu dieser Prognose. Auch in der *dlina*-Arbeitsgruppe waren wir, „like so many others“, am Beginn unserer Studien zur Netzwerkanalyse von Dramen der Überzeugung, bald einen neuen Blick auf die Geschichte der deutschsprachigen Dramatik vom Anfang der Aufklärung bis in die Frühe Moderne zu gewinnen. Und auch wir sind im Zuge unserer Analysen ins *Rabbit Hole* der „conceptual issues“ gefallen, die eine Eigendynamik entwickelt haben, die theoretische Konsequenzen haben könnte – Konsequenzen für die Weise, wie die *Digitale Literaturwissenschaft* Begriffsarbeit und -politik betreibt, Konsequenzen also nicht nur für die Verwendung, sondern auch für die Einführung und Definition, für die Positionierung und Profilierung von Begriffen, bei der die *Digitale Literaturwissenschaft* etwa vor der Frage steht, wie sie mit ihrem experimentellen, spielerischen, möglicherweise häretischen Potenzial umgehen soll.⁴

Worin könnte ein solches häretisches Potenzial der Begriffsarbeit in digital-empirischen Studien liegen? Auf den ersten Blick sicher in jener Forschungslogik, die auch andere empirische Wissenschaften prägt und die aus Sicht einer vornehmlich qualitativen Forschung als *bloßer* Reduktionismus, aus Sicht einer quantitativen Forschung hingegen als ein, sagen wir: *fröhlicher* Reduktionismus erscheinen muss. Gegenüber den traditionell „holistischen Phänomenerfahrungen

³Weitere Informationen auf dem Blog <https://dlina.github.io/und> im *GitHub Repo* <https://github.com/dlina>. Das *dlina*-Projekt wurde zuletzt weiterentwickelt und geht nun im *Programmable Corpora*-Projekt auf, das auf den *dracor*-Korpora und -Implementierungen beruht, wie sie von Frank Fischer, Carsten Milling, Mathias Göbel und Ingo Börner für <https://dracor.org> und Ivan Pozdniakov für die App *Shiny* <https://shiny.dracor.org/entwickelt> werden – derzeit jeweils noch im Alpha-Status (letzter Aufruf der Links 30.5.2019). Dazu siehe Frank Fischer/Ingo Börner/Mathias Göbel u. a., „Programmable Corpora. Die digitale Literaturwissenschaft zwischen Forschung und Infrastruktur am Beispiel von DraCor“, in: *DHd 2019. Digital Humanities. multimedial & multimodal. Konferenzabstracts*, Mainz/Frankfurt a. M., 194–197, <https://doi.org/10.5281/zenodo.2596094> (letzter Aufruf 30.5.2019).

⁴Dass dieses häretische Potenzial weit mehr betrifft als nur die Begriffe, sei hier lediglich erwähnt. An anderer Stelle haben Frank Fischer und ich erörtert, ob man sich aus einer praxeologischen Perspektive nicht – zumindest in Hinblick auf die Modellierung literarischer Texte nach den Maßgaben der formalen Netzwerkanalyse – von der Annahme verabschieden muss, das ‚epistemische Ding‘ (Hans-Jörg Rheinberger) der *Digitalen Literaturwissenschaft* sei identisch mit dem Wissenobjekt etwa der hermeneutischen Textinterpretation: Zu unterschiedlich sind die Praktiken, die die digitale Literaturwissenschaftler:in vollzieht, die Ausdrucksformen, die er wählt, die ‚technischen Dinge‘, mit denen sie ihren Gegenstand nicht nur umstellt, sondern erst zu einem ‚epistemischen Ding‘ macht. Siehe Peer Trilcke/Frank Fischer, „Literaturwissenschaft als Hackathon. Zur Praxeologie der Digitalen Literaturwissenschaft und zu ihren epistemischen Dingen“, in: Martin Huber/Sybille Krämer (Hg.), *Wie Digitalität die Geisteswissenschaften verändert. Neue Forschungsgegenstände und Methoden*, Sonderband der Zeitschrift für digitale Geisteswissenschaft 3 (27. Juni 2018), https://doi.org/10.17175/sb003_003 (letzter Aufruf 30.5.2019).

des Literaturwissenschaftlers⁵ sind die systematisch in ihrer Reichweite eingeschränkten Aussagen, die in der digital-empirischen Analyse erarbeitet werden, dürftig. Diese Dürftigkeit als Chance für eine Phänomenerkenntnis von bisher ungekannter empirischer Güte und bemerkenswerter Exaktheit zu begreifen, ist, denke ich, begründendes Moment der als häretisch empfundenen Fröhlichkeit zahlreicher Arbeiten aus der *Digitalen Literaturwissenschaft*.⁶

Für Moretti jedoch sind diese empirische Güte sowie die Präzision der Aussagen in der digital-empirischen Analyse nebensächlich. Die Provokation, die die ‚Computerisierung‘ sowie die damit einhergehende transparente und quantifizierende Operationalisierung von Begriffen für die Literaturwissenschaft darstellt, liege, so Moretti, nicht darin, dass bisher unscharfe Aussagen durch exakte Messungen präzisiert werden. „[I]f this is all measurement can do, then its role within literary study will only be a limited and ancillary one; making existing knowledge somewhat better, but not really different.“⁷

Bemerkenswert ist nun, dass Moretti darauf verzichtet, ein Argument anzuführen, das an dieser Stelle zu erwarten ist, weil Moretti selbst es in seinen *Distant-Reading*-Studien kultiviert hat: das Argument der *Large Scale*, des „great unread“.⁸ Denn die Operationalisierung von Begriffen im Sinne des *Measurement*-Paradigmas führt nicht nur zu einem neuen Grad von Exaktheit, sie ermöglicht auch die Ausweitung des Begriffsgebrauchs auf große, praktisch unlesbare Korpora. Während die bisher in der Literaturwissenschaft übliche Begriffsarbeit den Gebrauch der Begriffe an einen intransparenten und argumentationsaufwendigen hermeneutischen Prozess delegiert, lassen sich die nach dem *Measurement*-Paradigma operationalisierten Begriffe automatisiert anwenden. Dass der *begriffsintensionale* Reduktionismus der digital-empirischen Analyse insofern das Potenzial hat, die *begriffsextensionalen* Begrenztheiten der üblichen literaturwissenschaftlichen Praxis zu überwinden, ist dementsprechend ein verbreitetes Argument für die *Digitale Literaturwissenschaft*. Moretti sieht die „radicalization of our relationship to concepts“ jedoch weder darin, dass die digital-empirische Analyse die tradierten Begriffe formalisiert, reduziert, abbaut, also empirisch dekonstruiert (was ja tatsächlich das Verhältnis zu den Begriffen ändert, da ein wesentlicher Teil der Begriffsarbeit nun auf die Formulierung von

⁵Fotis Jannidis, „Digitale Literaturwissenschaft“, in: *Digitalität. Theorien und Praktiken des Digitalen in den Geisteswissenschaften* (4. Juli 2016), <https://digigeist.hypotheses.org/114> (letzter Aufruf 30.5.2019).

⁶Einschlägig hier die Kritik von Stephen Marche, „Literatur Is not Data. Against Digital Humanities“, in: *Los Angeles Review of Books* (28. Oktober 2012), <https://lareviewofbooks.org/article/literature-is-not-data-against-digital-humanities/> (letzter Aufruf 30.5.2019).

⁷Moretti (Anm. 2), 5.

⁸Mit Bezug auf Margaret Cohen bei Franco Moretti, „Conjectures on World Literature“, in: Ders., *Distant Reading*, London/New York 2013, 43–62, hier: 46.

strikten Regeln und Arbeitsschritten für den konkreten Gebrauch entfällt), noch darin, dass sie deren Anwendbarkeit durch Automatisierung ausweitet.

Worin die Provokation, die von der Begriffsarbeit der *Digitalen Literaturwissenschaft* ausgehen könnte, Moretti zufolge liegt, lassen die Beispiele erahnen, die er im sechsten *Pamphlet* anführt. So erwägt er unterschiedliche Operationalisierungen für den Begriff des Protagonisten, wobei seine Analysen zu einer Verschiebung des Fokus führen, der sich vom ‚Protagonisten‘ hin zu einem allgemeinen Konzept von Zentralität bewegt, was ihn zu einem Umbau in der literaturwissenschaftlichen Begriffshierarchie veranlasst („the ‚protagonist‘ [...] is only a special instance of the more general category of ‚centrality‘“⁹). In einem anderen Beispiel spielt Moretti Operationalisierungen des Begriffs der ‚tragischen Kollision‘ über die „most distinctive words“ durch, wobei er sich aufgrund der Ergebnisse zu einer Korrektur Hegels veranlasst sieht („this time, god forgive me, the mistake was Hegel’s“¹⁰).

In beiden Fällen ist es keineswegs so, dass Moretti bestehende Begriffe explizit verwirft, vielmehr wird die Operationalisierung zum Motor einer Begriffsdynamik, die zur Verschiebung der Begriffe oder, so im Fall der Zentralität, zur Einführung neuer, flankierender Begriffe führt.¹¹ Dabei wird die Operationalisierung als Motor dieser Begriffsdynamik unter anderem durch die Logik der Programme angetrieben, an die der Begriffsgebrauch im Zuge der Operationalisierung teilweise delegiert wird. Ins Spiel kommt damit, so Moretti, „the programming imagination“: „a form of thinking that fuses together the formulation and the operationalization of concepts, leaving them often half-implicit as concepts, while liberating their full force as algorithms.“¹² Zentralität, wie Moretti sie im sechsten *Pamphlet* diskutiert, ist so ein Fall: Als literaturwissenschaftlicher Begriff (noch) vage, diffus, als netzwerkanalytischer Algorithmus hingegen präzise bestimmt.

Morettis Überlegungen gewinnen an Kontur, wenn man zwei Szenarien der digital-empirischen Begriffsarbeit in der Literaturwissenschaft einander gegenüberstellt: Beim ersten Szenario steht zunächst ein tradiert Begriff, für den eine digital-empirische Operationalisierung entwickelt wird; dann wird geprüft, wie gut diese Operationalisierung den Begriff ‚abbildet‘ (etwa durch den Vergleich mit hermeneutischen Begriffsverwendungen und die Angabe von *Precision* und *Recall*), was gegebenenfalls zu einer Korrektur der Operationalisierung führt. Auch wenn eine vollständige ‚Abbildung‘ in den meisten Fällen unmöglich sein

⁹ Moretti (Anm. 2), 8.

¹⁰ Ebd., 12.

¹¹ Ein anderes Beispiel, das Moretti im zwölften *Pamphlet* diskutiert, ist der Begriff der *Loudness*, dazu siehe Holst Katsma, „Loudness in the Novell“, in: *Literary Lab Pamphlet* 7 (September 2014), <https://litlab.stanford.edu/LiteraryLabPamphlet7.pdf> (letzter Aufruf 30.5.2019).

¹² Moretti (Anm. 1), 6.

wird, ist das Ziel doch eine möglichst adäquate Operationalisierung. Im zweiten von Moretti umkreisten Szenario stehen zu Beginn zwar auch ein tradiert Begriff und dessen digital-empirische Operationalisierung. Dann jedoch übernimmt die Operationalisierung die Führung. Sie wird nicht durch die stete Rückkopplung an den tradierten Begriff justiert, ihre Eigendynamik wird freigesetzt, was auch bedeutet, der Logik der verwendeten Algorithmen (im engen Sinne von Programmen wie im weiten Sinne von definierten Arbeitsschritten) zu folgen. Dies führt zu einer Art *Concept Trouble*: etwa zu einer Verschiebung des Begriffs, zum Auftauchen neuer Begriffe oder zu einer Proliferation, zu einer Wucherung von Begriffen und zu einem Abdriften in andersartige terminologische Umwelten, in denen sich die eigene Arbeit zu bewegen, zu behaupten, zu erklären hat.

Die beiden Studien aus dem *dlna*-Projekt,¹³ die im Folgenden referiert werden, können zeigen, wie sich ein solches Abdriften vollzieht. Getrieben durch den methodischen Ansatz dieses Projekts, die formale Netzwerkanalyse, deren Zugriff auf die Objekte sich mit gegenstandsspezifischen Fragen und Begriffen vermengte, emergierten in diesen beiden Studien Begriffe, die in der Literaturwissenschaft bisher fremd sind. Wie das geschehen konnte, legen die beiden Studien dar, die in Abschn. 3 und 4 referiert und in Abschn. 5 kommentiert werden. Zuvor werden in Abschn. 2 einige Voraussetzungen dargelegt.

2 Modellierung und Preprocessing

Die Netzwerkanalyse beschreibt ihren Gegenstand (oder anders gesagt: sie konstituiert ihr ‚epistemisches Ding‘) als eine Struktur aus Elementen und Relationen (auch: Knoten/*Nodes* und Kanten/*Edges*). Sowohl Elemente als auch Relationen können mit quantitativen wie qualitativen Attributen versehen werden; Relationen können darüber hinaus gerichtet oder ungerichtet sein. Netzwerke lassen sich als Graphen visualisieren und mathematisch beschreiben.

¹³Die Studien wurden bereits auf *DH*-Konferenzen präsentiert, vgl. folgende (teilweise lediglich die Untersuchungsvorhaben skizzierenden) Abstracts: Für die erste Studie (Abschn. 3): Peer Trilcke/Frank Fischer/Mathias Göbel u. a., „Dramen als ‚Small Worlds‘? Netzwerkdaten zur Geschichte und Typologie deutschsprachiger Dramen 1730–1930“, in: *DHd 2016. Modellierung – Vernetzung – Visualisierung. Konferenzabstracts*, Leipzig 2016, 255–258, <http://dhd2016.de/boa.pdf> (letzter Aufruf 30.5.2019); Peer Trilcke/Frank Fischer/Mathias Göbel u. a., „Theatre Plays as ‚Small Worlds‘? Network Data on the History and Typology of German Drama, 1730–1930“, in: *Digital Humanities 2016. Conference Abstracts*, Kraków 2016, 385–387, <http://dh2016.adho.org/abstracts/360> (letzter Aufruf 30.5.2019). Für die zweite Studie (Abschn. 4): Peer Trilcke/Frank Fischer/Mathias Göbel, „Netzwerkdynamik, Plotanalyse. Zur Visualisierung und Berechnung der ‚progressiven Strukturierung‘ literarischer Texte“, in: *DHd 2017. Digitale Nachhaltigkeit. Konferenzabstracts*, Bern 2017, 175–180, http://www.dhd2017.ch/wp-content/uploads/2017/03/Abstractband_def3_M%C3%A4rz.pdf (letzter Aufruf 30.5.2019); Frank Fischer/Mathias Göbel/Dario Kampkaspar u. a., „Network Dynamics, Plot Analysis. Approaching the Progressive Structuration of Literary Texts“, in: *Digital Humanities 2017. Conference Abstracts*, Montréal 2017, 437–441, <https://dh2017.adho.org/abstracts/071/071.pdf> (letzter Aufruf 30.5.2019).

Gegenstand der Netzwerkanalyse in der *dlina*-Arbeitsgruppe sind dramatische Texte, derzeit aus der deutsch-, der französisch- und der russischsprachigen Literatur, wobei es im Folgenden um deutschsprachige Dramen gehen wird. Das deutschsprachige Korpus, das in den beiden Studien untersucht wurde und 465 Dramen aus der Zeit 1731 bis 1929 umfasst, haben wir auf Grundlage des *TextGrid Repository* erstellt,¹⁴ wobei wir die Daten aus diesem Repository noch einmal regelbasiert überarbeitet haben.¹⁵ Derzeit bauen wir das Korpus vollständig neu auf.¹⁶ Die im Folgenden referierten Ergebnisse beziehen sich jedoch noch auf das sog. *dlina-Corpus 15.07*.

Die Netzwerke, die wir aus den dramatischen Texten extrahiert haben, setzen als Elemente die Figuren eines Dramas, sofern diese Figuren innerhalb des Dramas einen Sprechakt vollziehen, d. h. sofern sie in der *TEI*-Struktur der Datei innerhalb eines `<speaker>`-Tags als Sprecher ausgewiesen werden. Als Relation zwischen zwei Figuren haben wir eine bestimmte Form der ‚Interaktion‘, die wir ‚szenische Kopräsenz‘ nennen, definiert: Eine solche ‚szenische Kopräsenz‘ und damit eine Relation zwischen zwei Figuren liegt demnach dann vor, wenn beide Figuren innerhalb eines durch die Struktur des Dramas vorgegebenen Segments (d. h. in der Regel: innerhalb einer Szene bzw. eines Auftritts; liegt keine Szenen- bzw. Auftritt-Unterteilung vor, wurde der Akt bzw. der Aufzug als Segment angesetzt) jeweils mindestens einen Sprechakt vollziehen.

Selbstverständlich gibt es ‚reichere‘ Optionen der Modellierung dramatischer Texte als Netzwerke. So wird bei unserer Operationalisierung von ‚Interaktion‘ nicht erfasst, ob Figuren tatsächlich *miteinander* sprechen. Auch die Ebene der Konfiguration wird von unserer szenen-, im Einzelfall sogar aktbasierten Segmentdefinition ignoriert. Schon die Art, wie der Untersuchungsgegenstand modelliert wird, ist also ein Beispiel für den fröhlichen Reduktionismus – mit all seinen Nachteilen und Vorteilen.

Präzise beschreiben ließen sich mit den Daten, die nach dieser Modellierung mithilfe des von Christopher Kittel und Frank Fischer entwickelten Tools *dramavis*¹⁷ erhoben wurden, einige basale Entwicklungen in der Dramengeschichte, denen

¹⁴Näheres erläutert der Blogpost: Frank Fischer/Peer Trilcke, *Introducing dlina Corpus 15.07 (Codename: Sydney)*. 20. Juni 2015, <https://dlina.github.io/Introducing-DLINA-Corpus-15-07-Codename-Sydney/> (letzter Aufruf 30.5.2019).

¹⁵Näheres zu diesen Preprocessing-Schritten, einschließlich der Editionsrichtlinien, erläutern folgende Blogposts: Dario Kampkaspar/Frank Fischer/Peer Trilcke, „Introducing Our ‚Zwischenformat‘“, in: *Digital Humanities and Literary Studies* (21. Juni 2015), <https://dlina.github.io/Introducing-Our-Zwischenformat/> (letzter Aufruf 30.5.2019); Dario Kampkaspar/Peer Trilcke, „Editing Rules“, in: *Digital Humanities and Literary Studies* (22. Juni 2015), <https://dlina.github.io/Editing-Rules/> (letzter Aufruf 30.5.2019).

¹⁶Vgl. die Informationen und Links zum *DraCor*-Projekt in Anm. 3.

¹⁷Christopher Kittel/Frank Fischer, „*dramavis* (v0.4)“, in: *GitHub Repository* (September 2017), <https://github.com/lehkost/dramavis> (letzter Aufruf 30.5.2019).

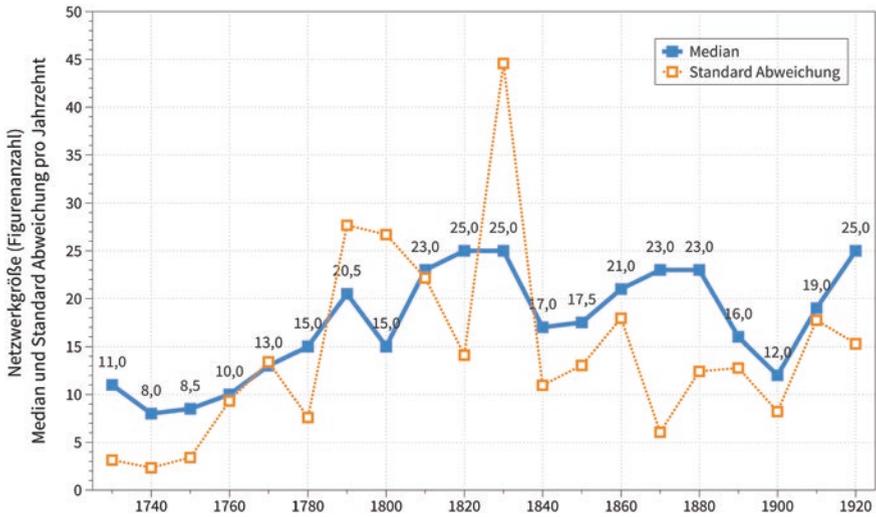


Abb. 1 Netzwerkgröße der Dramen im *dlina*-Korpus¹⁸

sich die ersten Analysen der Arbeitsgruppe widmeten. Berechnet wurde z. B. die Netzwerkgröße, also die Menge der Elemente in einem Netzwerk, d. h. die Figurenanzahl pro Drama, wobei in Abb. 1 jeweils für die Dramen einer Dekade der Median sowie die Standardabweichung dargestellt sind. Ebenfalls berechnet wurde der *Average Degree*: Dessen Grundlage ist der *Degree* jeder Figur (also hier die Anzahl der Relationen einer Figur zu anderen Figuren); der *Average Degree* eines Dramennetzwerks ist dann der Mittelwert der *Degrees* sämtlicher Figuren eines Dramas. Auch hier wurden Mediane und Standardabweichung pro Dekade gebildet (Abb. 2).

Die untersuchten Dramennetzwerke werden im Verlauf der Jahrhunderte nicht nur insgesamt umfangreicher (Abb. 1), die Figuren interagieren im Durchschnitt auch mit einer größeren Anzahl anderer Figuren (Abb. 2). Das hatten wir erwartet. Der Zeitraum, den das untersuchte Korpus abdeckt, ist nicht zuletzt eine Phase der umfassenden Modernisierung der Gesellschaft, einhergehend mit einem entsprechenden Komplexitätszuwachs sowie einer Ausdifferenzierung sozialer Beziehungen. Dass Literatur mit diesen sozialen Transformationsprozessen irgendwie interagiert (sie ‚abbildet‘, sie ‚widerspiegelt‘, sie ‚reflektiert‘, sich zu ihnen ‚homolog verhält‘ etc.), ist ein Basistheorem nicht nur von sozialgeschichtlich orientierten Literaturtheorien.

Auch die meisten Schwankungen in den Medianen waren wenig überraschend, so der Anstieg der Kurven in der zweiten Hälfte des 18. Jahrhunderts, in dem sich der Übergang vom Drama der Frühaufklärung u. a. zum sozialen Drama des Sturm und Drang abzeichnet; oder, in Hinblick auf die Netzwerkgröße, die ersichtlichen

¹⁸ Median und Standardabweichung pro Dekade.

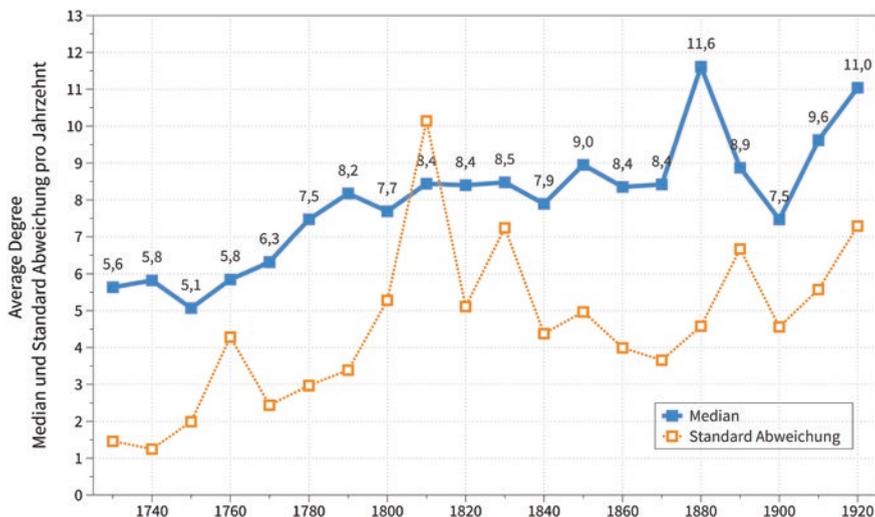


Abb. 2 Average Degree der Dramen im *dlna*-Korpus¹⁹

Spitzen in den 1810er, 1820er und 1830er Jahren, in denen sich die große Rolle des *Historischen Dramas* zeigt; oder das Absacken der Netzwerkgröße in und nach den 1840er Jahren, ein erwartbarer Effekt des Rückzugs ins Private in Folge der gescheiterten Revolution; sowie das in beiden Diagrammen zu beobachtende, erneute Absacken der Kurven in den 1890er Jahren, in denen die naturalistische Dramatik sowie andere, avantgardistische Experimente für kleine Off-Bühnen zu verorten sind.

Die Kurven der Standardabweichung differenzieren dieses Bild etwas: Das Ansteigen und Abfallen der Kurven deutet darauf hin, dass es wiederholt zu Phasen der Ausdifferenzierung der Dramen in umfangreichere und beschränktere Typen kommt, auf die dann jeweils eine Phase der Homogenisierung folgt, so etwa – mit Blick auf die Netzwerkgröße – in den Jahren 1750 bis 1780, in denen der Typus des offenen Dramas zunächst langsam an die Seite des Dramas der Aufklärung tritt, das dann später verschwindet. Ähnliche Entwicklungen lassen sich beim *Average Degree* in den 1760 und 1770er Jahren oder 1810er und 1820er Jahren ausmachen. Starke Ausschläge, wie der bei der Netzwerkgröße in den 1830er Jahren, führen dabei das Auftreten von Extremtypen, in diesem Fall der Grabbe-Dramen *Napoleon* und *Hannibal* sowie Goethes *Faust II*, vor Augen (und machen zudem die dann doch beschränkte Größe des untersuchten Korpus deutlich).

Zu den Diagrammen lassen sich ohne Weiteres Literaturgeschichten erzählen. Diese sind allerdings bekannt; das Wissen besteht bereits und wird auf die Daten

¹⁹Median und Standardabweichung pro Dekade.

angewendet.²⁰ Ganz überraschungsfrei ist das nicht, in der Regel aber lässt sich, was zunächst als Überraschung erscheint, durch Eigenarten der Korpuszusammensetzung erklären,²¹ womit sich wiederum das ‚alte‘ Wissen durchsetzt. Das bedeutet nicht, dass solche datengestützten Wiedererzählungen bekannter Literaturgeschichten unbefriedigend sind. Und doch gilt hier der oben bereits angeführte Schluss Morettis: „[I]f this is all measurement can do, then its role within literary study will only be a limited and ancillary one; making existing knowledge somewhat better, but not really different.“²²

3 Small Worlds

Eine andere Dynamik gewannen die Arbeiten im *dlina*-Projekt erst etwas später, als wir die strukturelle Komposition der Dramen in den Blick nahmen. Diese war bereits einmal in den Vorstudien zum *dlina*-Projekt adressiert worden, dort ausgehend von der in der Literaturwissenschaft etablierten typologischen Unterscheidung in eine offene und geschlossene Form des Dramas, wie sie Volker Klotz in den 1960er Jahren entwickelt hat.²³ Klotz' typologische Begriffe sind multi-dimensional angelegt, sie umfassen u. a. die Aspekte ‚Handlung‘, ‚Zeit‘, ‚Raum‘ usw. Daneben thematisiert Klotz aber auch den Aspekt ‚Komposition‘, dessen Bestimmungen sich gut für die quantitative Operationalisierung eignen. Ein Beispiel für eine dieser Operationalisierungen sei hier angeführt:

Geht man davon aus, dass im offenen Drama eine Figur „[f]ast in jeder neuen Szene [...] anderen Personen gegenüber[steht]“, von diesen anderen Personen bzw. Figuren jedoch viele „nur in einer einzigen Szene auf[treten]“, dann ist zu erwarten, dass eine oder wenige Figuren (z. B. das von Klotz so genannte ‚zentrale Ich‘) in relativ vielen Beziehungen stehen, mithin einen hohen Degree aufweisen, wohingegen zahlreiche Figuren (die von Klotz so genannten ‚atmosphärischen Figuren‘) in relativ wenigen Beziehungen stehen, mithin einen niedrigen Degree aufweisen. Diese Streuung der Degrees im offenen Drama müsste sich dabei in einer relativ hohen Standardabweichung (Standard Deviation, SD) niederschlagen.²⁴

²⁰ Gerade bei der Analyse großer Korpora ist dieses Vorgehen des ‚bekannte Geschichten zu neuen Daten‘-Erzählens recht weit verbreitet, vgl. z. B. jüngst Mark Algee-Hewitt, „Distributed Character: Quantitative Models of the English Stage“, 1500–1920, in: *Digital Humanities 2017. Conference Abstracts*, Montréal 2017, 119–121, <https://dh2017.adho.org/abstracts/103/103.pdf> (letzter Aufruf 1.5.2018).

²¹ So etwa, um nur ein Beispiel zu nennen, der zunächst bemerkenswerte Ausschlag beim *Average Degree* in den 1880er Jahren (Abb. 2), der u. a. auf zwei Dramen von Karl Bleibtreu (*Weltgericht*, 1888, und *Ein Faust der That*, 1889) zurückzuführen ist, die allein eine Aktunterteilung, keine Szenenunterteilung aufweisen, was bei unserem Ansatz in der Regel einen höheren *Average Degree* produziert. Ohne diese beiden Dramen liegt der Median bei 9,9.

²² Moretti (Anm. 2), 5.

²³ Volker Klotz, *Geschlossene und offene Form im Drama*, München ³1968.

²⁴ Peer Trilcke, „Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft“, in: Philip Ajouri/Katja Mellmann/Christoph Rauen (Hg.), *Empirie in der Literaturwissenschaft*, Münster 2013, 201–247, hier: 225 (Die Zitate beziehen sich auf Klotz [Anm. 23], 154).

Neben diesem Indikator wurden drei weitere formuliert und schließlich exemplarisch anhand von vier Dramen auf ihre Anwendbarkeit überprüft. Goethes Drama *Götz von Berlichingen* erwies sich, ganz im Sinne von Klotz, als Prototyp der offenen Form (die Standardabweichung der Degrees war hier z. B. 7,14), wohingegen sich Goethes *Iphigenie auf Tauris* als Prototyp der geschlossenen Form zeigte (Standardabweichung der Degrees: 0).²⁵ Erneut war es wie erwartet. Erneut war das Wissen im Grunde unverändert geblieben.

Doch die von Klotz ausgehende typologische Analyse war ein Einfallstor für eine Analyseperspektive auf die Dramen, die im Verlauf der weiteren Untersuchung etwas anderes erkennbar werden ließ. Die in der oben angeführten Operationalisierung verwendete Analyse der „Streuung der Degrees“, so zeigte sich, war in der Netzwerkanalyse ein übliches Verfahren. Zur Anwendung kam sie auch dort im Kontext der Typologisierung von Netzwerken: als *Node Degree Distribution (NDD)*. Eine bestimmte Form der NDD, konkret eine NDD mit *Power-Law-Verteilung*, gilt als Kriterium für sog. *Scale-Free-Netzwerke*.²⁶ *Scale-Free-Netzwerke* werden als eine spezifische Variante von *Small-World-Netzwerken* begriffen. Und mit diesem in der Netzwerktheorie seit einem Artikel von Duncan J. Watts und Steven H. Strogatz etablierten Begriff begann das Abdriften des *dllina*-Projekts;²⁷ die im Zuge der Studie zu Klotz' Typen bereits erfolgte Operationalisierung lockte gewissermaßen in ein *Rabbit Hole*, in dem eine ganz neue Begriffswelt wartete.

Im Zentrum dieser Begriffswelt steht die *Small World*.²⁸ Auf ihr Vorkommen hin (sowie auf das der spezifischen Variante, des *Scale-Free-Netzwerks*), so war das Ziel der Studie, sollte das *dllina*-Korpus untersucht werden, was auch deshalb sinnvoll erschien, weil dieser Typus von Watts und Strogatz als „widespread in biological, social and man-made systems“²⁹ – und warum also nicht auch in der Literatur – charakterisiert wurde.³⁰ *Small-World-Netzwerke* (Abb. 3b) sind in Relation zu zwei anderen Netzwerktypen definiert: das reguläre (Abb. 3a) und

²⁵ Siehe ebd., 235.

²⁶ Siehe Réka Albert/Albert-László Barabási, „Statistical Mechanics of Complex Networks“, in: *Reviews of Modern Physics* 74 (2002), 47–97.

²⁷ Duncan J. Watts/Steven H. Strogatz, „Collective Dynamics of ‚Small World‘ Networks“, in: *Nature* 393 (1998), 440–442.

²⁸ Erste Anwendungen des *Small-World*-Begriffs auf Dramen finden sich bei James Stiller/Daniel Nettle/Robin I. M. Dunbar, „The Small World of Shakespeare's Plays“, in: *Human Nature* 14 (2003), 397–408, und James Stiller/Matthew Hudson, „Weak Links and Scene Cliques Within the Small World of Shakespeare“, in: *Journal of Cultural and Evolutionary Psychology* 3 (2005), 57–73.

²⁹ Watts/Strogatz (Anm. 27), 442.

³⁰ Darüber hinaus lag unserer *Small-World*-Analyse der Gedanke zugrunde, dass *Small Worlds* u. a. bestimmte, insbesondere im Zuge der Modernisierung emergierende Typen sozialer Formationen beschreiben. Dramen aber begreifen wir, wie jeden diegetischen literarischen Text, stets auch als kontextsensible ästhetische Modelle solcher Formationen. Es lag also nahe, zu prüfen, ob sich in unserem Korpus Texte identifizieren lassen, die als *Reflexionen* von sozialen *Small-World*-Formationen interpretiert werden könnten.

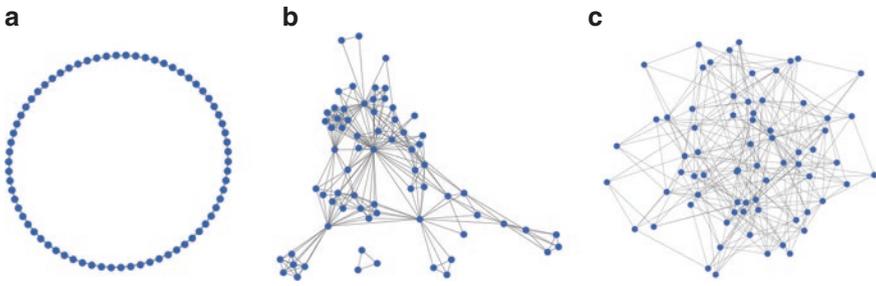


Abb. 3 a. Reguläres Netzwerk (links) – Small World (Mitte) – Random-Netzwerk (rechts) b. Small World c. Random-Netzwerk

Tab. 1 Exemplarische Werte (C und APL) für ein reguläres Netzwerk (links), ein *Small-World*-Netzwerk (Mitte) und ein *Random*-Netzwerk (rechts)

Clustering Coefficient (C)		
0,600	0,852	0,131
Average Path Length (APL)		
6,261	2,346	2,258

das *Random*-Netzwerk (Abb. 3c). Charakteristisch für ein *Small-World*-Netzwerk sind zwei Eigenarten: Erstens gibt es in *Small Worlds* kleinere, dicht verbundene Bereiche, sogenannte *Cliquen* oder *Cluster* (Abb. 4). Zweitens sind diese Cluster untereinander nur selten direkt verbunden, sondern zumeist vermittelt durch einige sehr zentrale Knoten, sogenannte *Netzwerkhubs* (Abb. 5).

Die Eigenarten lassen sich quantifizieren. Watts und Strogatz definieren *Small Worlds* als Klasse von Netzwerken, die „highly clustered, like regular lattices“ sind (und damit stark von *Random*-Netzwerken abweichen), „yet have small characteristic path lengths, like random graphs“.³¹ Die Werte zu Abb. 3a bis 3c veranschaulichen dies (s. Tab. 1). Die beiden Kriterien wurden für die Analyse folgendermaßen formuliert:

- Kriterium 1: Der *Clustering Coefficient (C)* des beobachteten Netzwerks, also des jeweiligen Dramen-Netzwerks, ist signifikant höher als der C eines entsprechenden *Random*-Netzwerks.
- Kriterium 2: Die *Average Path Length (APL)* des beobachteten Netzwerks weicht nicht signifikant ab von der APL eines entsprechenden *Random*-Netzwerks.

Um auch nach dem spezielleren Typus der *Scale-Free*-Netzwerke zu fragen, wurde zudem ein drittes Kriterium formuliert, das zusätzlich eine *Power-Law*-Verteilung der NDD verlangt, die als *Differentia Specifica* dieses Netzwerk-Typs gilt.³² Vereinfacht

³¹ Watts/Strogatz (Anm. 27), 440.

³² Albert/Barabási (Anm. 26), 71.

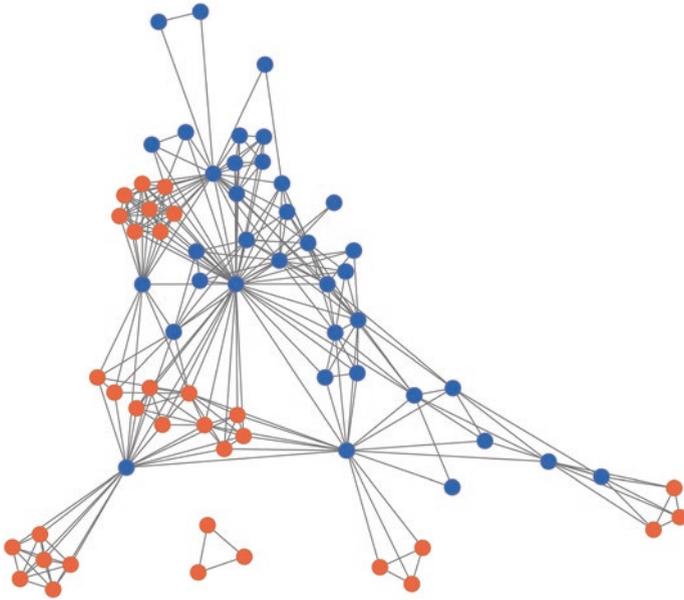


Abb. 4 Beispiel für ein Charakteristikum von *Small-World*-Netzwerken – *Cliquen*

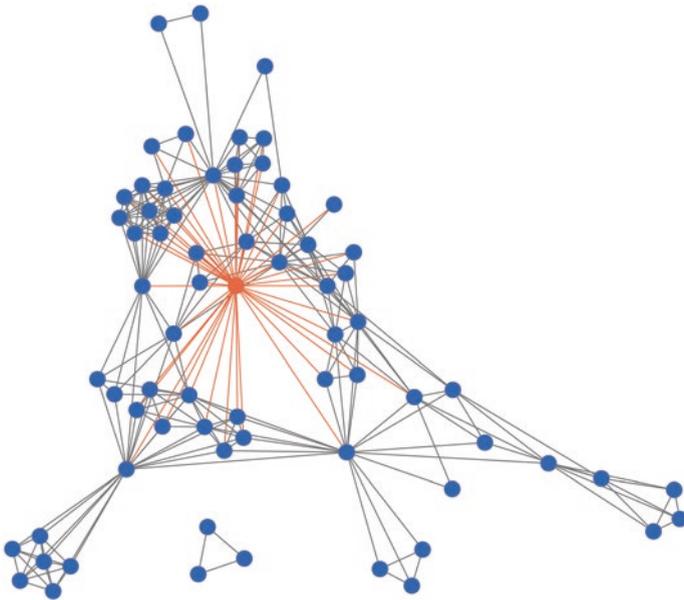


Abb. 5 Beispiel für ein Charakteristikum von *Small-World*-Netzwerken – zentrale Knoten (*Hubs*)

bedeutet eine *Power-Law*-Verteilung dabei, dass es sehr wenige Figuren mit einem sehr hohen *Degree* – also sehr wenige sehr gut vernetzte Figuren – gibt und zugleich sehr viele Figuren mit einem sehr geringen *Degree*. Erfüllt war dieses Kriterium dann, wenn

- Kriterium 3: die *Node-Degree*-Distribution sich am besten (im Sinne des höchsten Bestimmtheitsmaßes R^2) mit einer *Power Law*-Regression beschreiben ließ.

Im Folgenden wurden das *dolina*-Korpus auf diese Kriterien hin überprüft. Zunächst wurde jedes der 465 Dramen-Netzwerke randomisiert, wobei – mittels einer Implementierung im *dramavis*-Tool – zu jedem Netzwerk nach dem *Erdős-Rényi*-Modell ($G(n, M)$ ‘*Model*’) 1000 *Random*-Netzwerke berechnet und dann die Mittelwerte für CC und für APL dieser *Random*-Netzwerke erhoben wurden. Damit waren die Tests auf die Kriterien 1 und 2 möglich. Dafür wurde jeweils der Quotient aus dem CC bzw. dem APL des beobachteten Dramen-Netzwerks und des Mittelwerts für das tausendfach randomisierte Netzwerk berechnet. Daraufhin wurden einfache Signifikanztests durchgeführt: Zunächst wurden diejenigen Dramen-Netzwerke ausgewählt, deren CC signifikant von den CC des randomisierten Netzwerks abwichen, bei denen der berechnete Quotient also bemerkenswert hoch war (Grenze für die Signifikanz war ‚Mittelwert + $2 \times$ Standardabweichung‘). Im Ergebnis erfüllten 23 Dramen aus unserem Korpus dieses Kriterium 1 (vgl. Tab. 2). Im anschließenden Schritt wurden aus diesen 23 Dramen all jene aussortiert, bei denen eine signifikante Abweichung von der APL der randomisierten Netzwerke vorlag (Grenze für die Signifikanz war ‚Mittelwert +/– $2 \times$ Standardabweichung‘)³³ – sechs Dramen schieden infolgedessen aus, sodass 17 Dramen übrig blieben, die sowohl Kriterium 1 als auch Kriterium 2 erfüllten und insofern als *Small Worlds* gelten können (vgl. Tab. 2). In einem letzten Schritt erfolgte die Überprüfung dieser 17 Dramen auf das *Scale-Free*-Kriterium 3. Dabei erwies sich, dass sich lediglich bei fünf Dramen die NDD am besten mit einer *Power-Law*-Verteilung beschreiben lässt (vgl. Tab. 2).

Im Grunde war unsere Analyse damit erfolgreich beendet. Es ließen sich tatsächlich sowohl *Small-World*-Netzwerke als auch die spezielleren *Scale-Free*-Netzwerke in unserem Dramen-Korpus finden. Und es war keineswegs so, dass wir damit lediglich Dramen mit einem besonders großen Umfang des Personals, also mit vielen Figuren, identifiziert hatten, wie ein Blick auf Abb. 6 und 7 zeigt, bei denen jedes Kreuz für ein Drama steht: mit dem Erscheinungsdatum auf der Abszisse und der Netzwerkgröße (Figurenanzahl) auf der Ordinate. Auch wenn,

³³Da bei der APL eine annähernde Normalverteilung vorlag (anders als beim Quotienten aus CC des beobachteten Netzwerks und CC des randomisierten Netzwerks, wo eine exponentielle Verteilung vorlag), wurde hier +/– Standardabweichung für die Berechnung der Signifikanz ausgewählt.

Tab. 2 Ergebnisse des Tests auf die drei Kriterien

Autor/in	Titel	Jahr	Kriterium 1	Kriterium 2	Kriterium 3
Goethe	<i>Götz von Berlichingen</i>	1773	x	x	x
Soden	<i>Doktor Faustus</i>	1797	x	x	x
Arnim	<i>Jerusalem</i>	1811	x	x	x
Raimund	<i>Der Barometermacher</i>	1823	x	x	x
Nestroy	<i>Der böse Geist</i>	1833	x	x	x
Tieck	<i>Prinz Zerbino</i>	1799	x	x	
Schiller	<i>Die Jungfrau von Orleans</i>	1801	x	x	
Kleist	<i>Die Hermannsschlacht</i>	1808	x	x	
Arnim	<i>Halle</i>	1811	x	x	
Gleich	<i>Der Ehetüfel auf Reisen</i>	1821	x	x	
Voß	<i>Faust</i>	1823	x	x	
Raimund	<i>Die unheilbringende Zauberkrone</i>	1829	x	x	
Birch-Pfeiffer	<i>Die Walpurgisnacht</i>	1830	x	x	
Immermann	<i>Andreas Hofer</i>	1835	x	x	
Vischer	<i>Faust</i>	1835	x	x	
Panizza	<i>Nero</i>	1898	x	x	
Avenarius	<i>Faust</i>	1919	x	x	
Beer	<i>Struensee</i>	1828	x		
Grabbe	<i>Napoleon</i>	1831	x		
Goethe	<i>Faust II</i>	1832	x		
Lassalle	<i>Franz von Sickingen</i>	1859	x		
Schäfer	<i>Faustine</i>	1898	x		
Hofmannsthal	<i>Die Frau im Fenster</i>	1898	x		

was wir erwartet hatten, *Small-World*- und *Scale-Free*-Strukturen gehäuft bei größeren Dramen auftreten, scheinen sie doch eine davon unabhängige Eigenart bestimmter Dramen zu sein, deren Erscheinen im Fall der *Scale-Free*-Netzwerke auch noch historisch auf die Goethe-Zeit beschränkt ist (vgl. Abb. 7). Darüber hinaus identifizierten die Analysen auffällig viele *Faust*-Dramen (4 von 17) sowie ebenso viele Zauberspiele bzw. Zauberpossen (Nestroy, Gleich, die beiden Raimund-Dramen) als *Small Worlds*. Allerdings erschöpfte sich unsere Klassifizierung nicht in diesen beiden genreartigen Textgruppen. Was aber hatten wir dann eigentlich gemessen, was für einen Dramen-Typus hatten wir identifiziert?

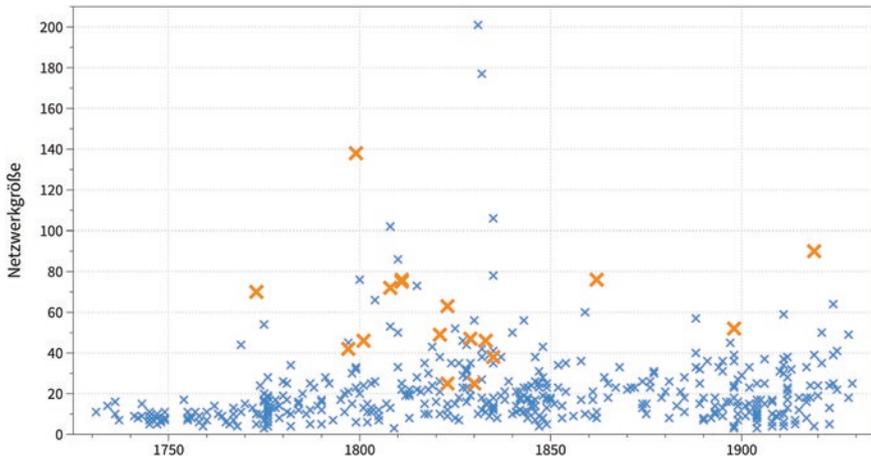


Abb. 6 17 *Small-World*-Netzwerke im *dlna*-Korpus

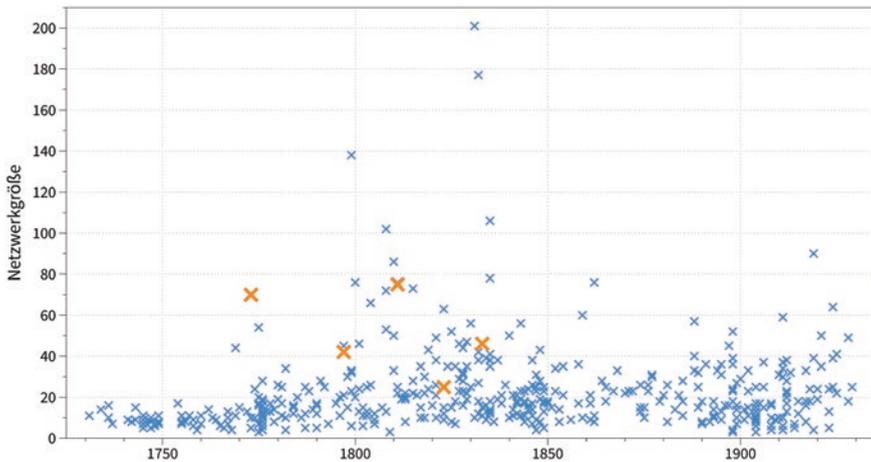


Abb. 7 5 *Scale-Free*-Netzwerke im *dlna*-Korpus

Zu dem *Concept Trouble*, der sich bei der Reflexion über diese Frage einstellte, gehört, dass wir immer wieder zurückgeworfen wurden auf die Einsicht, dass es eben *dramatische Small Worlds* sind, die wir identifiziert hatten. Sicher, die Struktur eines *Small-World*-Dramas lässt sich mit tradierten dramenästhetischen Begriffen, wie dem ‚zentralen Ich‘,³⁴ in Verbindung bringen (vgl. noch einmal Abb. 6); auch finden sich in ihr weitere Aspekte der dramatischen Tektonik der

³⁴ Klotz (Anm. 23), 106–109.

offenen Form, die Klotz beschrieben hat, etwa die „Streuung des Geschehens“, ³⁵ wie sie sich in den Clustern der *Small Worlds* abzeichnet. Doch geht die dramatische *Small World* in diesen tradierten Begriffen und Beschreibungen nicht auf. Zum unvertrauten Charakter dessen, was mit dem Begriff beobachtet wird, trägt zudem bei, dass die dramatische *Small World* ein Konstrukt ist, das auf der spezifischen Modellierung der Netzwerke über die szenische Kopräsenz basiert. Der verwendete *Small-World*-Begriff bezeichnet also keineswegs etwa eine bestimmte Art der ‚objektiven Sozialstruktur‘ in der dargestellten Welt, sondern eine Weise der ästhetischen Inszenierung von Figureninteraktion auf der Bühne. In die hier analysierte Struktur der Dramen-Netzwerke sind, anders gesagt, Effekte der dramatischen Komposition – z. B. auch der Handlungsführung, die veranlasst, dass eine Figur mit dieser oder jener anderen Figur zusammentrifft – eingeflossen.

Diese Reflektionen machen deutlich, dass es schwerfällt, den Begriff der dramatischen *Small World* und also die Ergebnisse unserer Analysen in tradiertes literaturwissenschaftliches Wissen zu integrieren, zu unvertraut sind die Gegenstandsmodellierung und die Begriffsoperationalisierung, die zugrunde gelegt wurden. Entstanden sind stattdessen zunächst einmal neue Informationen über literarische Texte, von denen unsicher ist, ob ihnen der Status *literaturwissenschaftlichen Wissens* zugeschrieben werden kann, wobei eine solche Zuschreibung letztlich Mechanismen der disziplinären Anerkennung und Anschlusskommunikation voraussetzen würde. Das provokative Potenzial jenes *Concept Troubles*, den man bei Moretti angedeutet findet, liegt aber womöglich darin, die Frage, ob und wie man sich noch im disziplinären Horizont der Literaturwissenschaft bewegt, für den Moment zu suspendieren – und sich von der Eigenlogik der Operationalisierung, der Programme und der Methoden noch weiter treiben zu lassen, *ohne* die Reflexion über die Beziehung zu den tradierten Konzepten aufzugeben.

Im Fall der *Small World*-Studie bedeutete das, dass wir in einem weiteren Schritt genauer auf die *Node Degree Distribution* blickten, die in das Kriterium 3 eingeflossen war. Im Rahmen der Studie war ein spezifischer Typus dieser NDD identifiziert worden: die *Power-Law*-Verteilung, bei der es wenige Figuren gibt, die mit sehr vielen anderen Figuren szenisch kopräsent sind, und sehr viele Figuren, die mit wenigen anderen Figuren szenisch kopräsent sind, so wie bei Goethes *Götz* der Fall (Abb. 8a und 8b). Aber was für andere *Degree*-Verteilungen gibt es? Was für andere, so könnte man auch fragen, *Typen der quantitativen Strukturierung des Personals* lassen sich in den Dramen des *dina*-Korpus finden?

Die Abb. 10a bis 13 zeigen Graphen und NDD-Diagramme von drei weiteren Typen der quantitativen Strukturierung des Personals, die sich – nach der Funktion der Regressionslinie, die die Verteilung mit dem höchsten Bestimmtheitsmaß beschreibt – als *quadratische* Verteilung (Abb. 9a und 9b), als *polynomische (4. Grad)* bzw. *quartische* Verteilung (Abb. 10a und 10b) und als *umgekehrte Power-Law*-Verteilung (Abb. 11a und 11b) bezeichnen lassen.

³⁵ Ebd., 103.

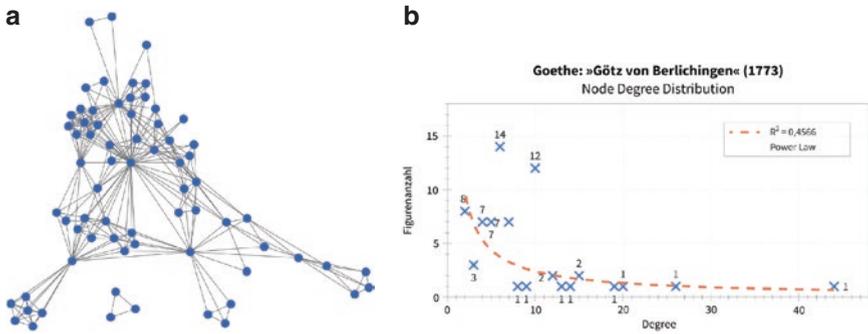


Abb. 8 a. Graph zu Goethe: *Götz* (1773) b. NDD-Diagramm zu Goethe: *Götz* (1773)

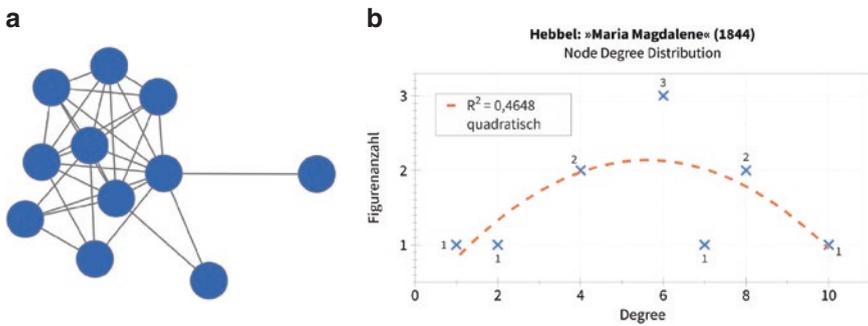


Abb. 9 a. Graph zu Hebbel: *Maria Magdalene* (1844) b. NDD-Diagramm zu Hebbel: *Maria Magdalene* (1844)

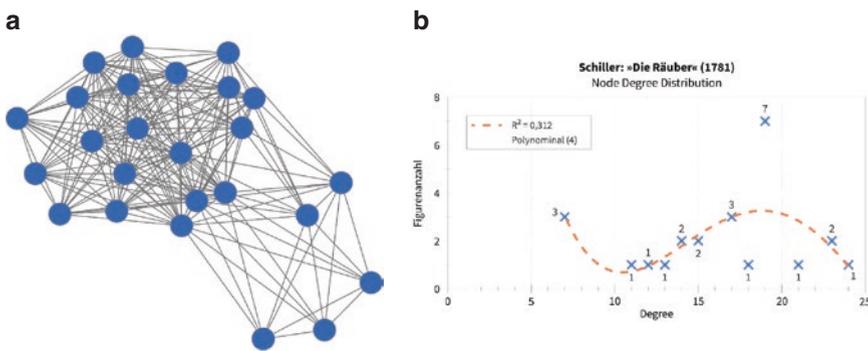


Abb. 10 a. Graph zu Schiller: *Die Räuber* (1781) b. NDD-Diagramm zu Schiller: *Die Räuber* (1781)

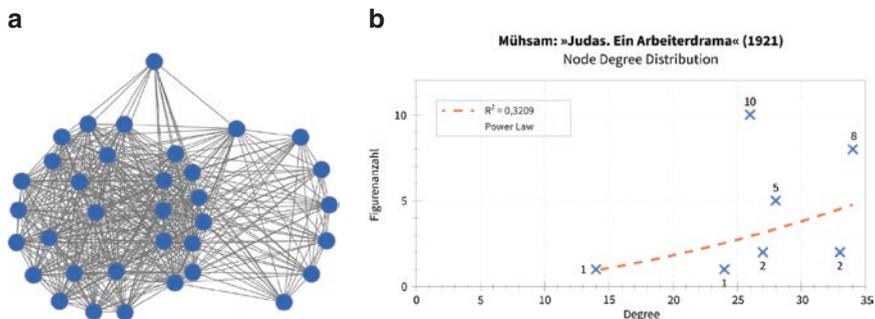


Abb. 11 a. Graph zu Mühsam: *Judas* (1921) b. NDD-Diagramm zu Mühsam: *Judas* (1921)

Gegenüber dem *Small-World*-Konzept sind diese Verteilungstypen deutlich schlichter definiert; und was sie zeigen, ist vermutlich auch deutlich ‚griffiger‘ für den literaturwissenschaftlichen Diskurs, so etwa wenn man das Drama des *großen Individuums*,³⁶ also Goethes *Götz* (Abb. 8a und 8b), und dessen *aristokratische Struktur* mit dem Drama der *Masse* und dessen, wie man sagen könnte: *kommunistischer Struktur* vergleicht, wie sie sich in Erich Mühsams ‚Arbeiterdrama‘ *Judas* zeigt (Abb. 11a und 11b). Zudem erweist sich die Analyse der *Degree*-Verteilung als deutlich kompatibler zu vorliegender Fachterminologie, insbesondere zu Manfred Pfisters Analysekatgorie der „quantitativen Dominanzrelation“,³⁷ in die sie einfließen könnte. Und vielleicht bedarf die Idee der *Small World* ja einer solchen Zerlegung in ihre Bestandteile bzw. Kriterien, um wiederum literaturwissenschaftlich anschlussfähig zu werden, jedenfalls im Sinne eines fröhlichen Reduktionismus, aus dem nicht unbedingt neue, aber sicher quantitativ-empirische Begriffe der Haupt-, der Neben-, der Episoden-Figuren usw. hervorgehen könnten.³⁹ Das Irritationspotenzial des *Small-World*-Begriffs, der den Diskurs auf befremdende Weise öffnet für eine allgemeine Strukturphänomenologie anthropogener Netzwerke, wäre damit allerdings vom Tisch.

³⁶ So eine typische Formulierung mit Blick auf die Sturm-und-Drang-Dramatik, vgl. explizit zu Goethes *Götz* etwa Andreas Huyssen, *Das Drama des Sturm und Drang. Kommentar zu einer Epoche*, München 1980, 154.

³⁷ Manfred Pfister, *Das Drama. Theorie und Analyse*, München 1997, 226 f.

³⁹ Siehe dazu u. a. Frank Fischer/Peer Trilcke/Christopher Kittel u. a., „To catch a protagonist: Quantitative dominance relations in german-language drama (1730–1930)“, in: *Digital Humanities 2018. Conference Abstracts, El Colegio de México & Universidad Nacional Autónoma de México & Red de Humanidades Digitales* 2018, 193–201, https://dh2018.adho.org/wp-content/uploads/2018/06/dh2018_abstracts.pdf (letzter Aufruf 30.5.2019).

4 Change Rates

Die Betrachtung von literarischen Texten als Netzwerke hat, das zeigen etwa auch vergleichsweise unbefriedigende Versuche der Genreklassifikation mittels Netzwerkmaßen,⁴¹ etwas Kontraintuitives, was nicht zuletzt an der ‚strukturalistischen Intuition‘⁴² der Netzwerkanalyse und dem damit einhergehenden Zug zur formalen Abstraktion liegt, der den Gegenstand der Analyse stärker hermeneutisch verfremdet als bedeutungsbezogene Verfahren wie die *Stylometrie* oder das *Topic Modeling*. Auch deshalb spitzt sich das provokative Potenzial der digital-empirischen Begriffsarbeit bei Verfahren der Netzwerkanalyse zu: Sind die Berechnungen etwa des *Clustering Coefficients* oder der *Average Path Length*, selbst bei entsprechender Toleranz gegenüber der Quantifizierung, wirklich angemessene Operationen angesichts ästhetischer Artefakte?

Diese offene Frage ist ein Motor jener Pendelbewegung im *dlna*-Projekt, die uns nach dem vorläufigen Ende der experimentellen *Small-World*-Studie zurück zu einem reichen literaturwissenschaftlichen Begriff brachte – und von dort, dem Sog der netzwerkanalytischen Eigenlogik ebenso folgend wie der ‚programming imagination‘, wieder ins *Rabbit Hole* der Operationalisierung fallen ließ.

Dass die Netzwerkanalyse das (provokative) Potenzial bergen könnte, einen neuen Ansatz zur *Plot Analysis* zu offerieren, hatte Moretti bereits im Titel seines zweiten Pamphlets angekündigt, um dann das Scheitern an diesem Anspruch effektiv zu inszenieren.⁴³ Ein Grund für dieses Scheitern schien uns, als wir uns der Frage nach der *Plot Analysis* zuwandten, offensichtlich: Moretti hatte, wie auch wir bis dahin, statische Netzwerke betrachtet. *Plot* jedoch, wie immer man diesen Begriff konkret versteht, hat eine temporale Dimension: ‚the repeated attempts to redefine parameters of plot reflect both the centrality and the complexity of the *temporal dimension* of narrative‘, heißt es bei Hilary Dannenberg,⁴⁴ und Karin Kukkonen versteht *Plot* als ein Konzept zur Beschreibung der ‚progressive structuration‘ von Texten.⁴⁵ Nicht Struktur also, sondern Strukturierung, nicht statische Netzwerke, sondern dynamische.

⁴¹ Siehe Lena Hettinger/Martin Becker/Isabella Reger u. a., ‚Genre classification on German novels‘, in: *Proceedings of the 12th International Workshop on Text-based Information Retrieval*, Valencia 2015, <http://www.uni-weimar.de/medien/webis/events/tir-15/tir15-papers-final/Hettinger2015-tir-paper.pdf> (letzter Aufruf 30.5.2019).

⁴² Linton C. Freeman, *The Development of Social Network Analysis. A Study in the Sociology of Science*, North Charleston 2004, 4.

⁴³ Franco Moretti, ‚Network Theory, Plot Analysis‘, in: *Literary Lab Pamphlet 2* (Mai 2011), <http://litlab.stanford.edu/LiteraryLabPamphlet2.pdf> (letzter Aufruf 30.5.2019).

⁴⁴ Hilary P. Dannenberg, ‚Plot‘, in: David Herman/Manfred Jahn/Marie-Laure Ryan (Hg.), *The Routledge Encyclopedia of Narrative Theory*, London 2005, 435–439, hier: 435.

⁴⁵ Karin Kukkonen, ‚Plot‘, in: Peter Hühn/Jan Christoph Meister/John Pier/Wolf Schmid (Hg.), *the living handbook of narratology* (24. März 2014), § 4, <http://www.lhn.uni-hamburg.de/article/plot> (letzter Aufruf 30.5.2019).

Eine solche Analyse dynamischer Netzwerk – *Dynamic Network Analysis* – hat bereits eine Tradition, deren Anfänge unter anderem in der Analyse von Terrornetzwerken liegen.⁴⁴ Inspiriert von diesen Analysen und den dort implementierten Algorithmen begannen wir, für die Dramen in unserem Korpus zahlreiche Werte zu berechnen: den *Central Character Entry Index*, den *All-in Index*, das *Final Scene Size Measure* oder die *Drama Change Rate*. Die meisten Werte waren dabei von einfachen literaturwissenschaftlichen Fragen abgeleitet: Wann tritt die Figur mit den höchsten Zentralitätswerten das erste Mal auf (*Central Character Entry Index*)? Wann war jede Figur mindestens einmal auf der Bühne (*All-in Index*)? Wie viele Figuren des Personals treten in der letzten Szene auf (*Final Scene Size Measure*)? Die *Drama Change Rate* hingegen war rein netzwerkanalytisch, wenn nicht algorithmisch inspiriert. Sie war es, die im Folgenden eine besondere Dynamik entwickelte.

Die *Drama Change Rate*, wie wir sie netzwerkanalytisch definieren, ist ein Maß für die Veränderung der Teilmenge des Personals, das sich in den vorgegebenen Segmenten eines Dramas (hier in der Regel Szenen bzw. Auftritte) auf der Bühne befindet. Grundlage für die Berechnung der *Drama Change Rate* ist die *Segment Change Rate*, die beschreibt, wie sich die Menge der auf der Bühne präsenten Teilmenge des Personals zwischen zwei aufeinanderfolgenden Segmenten verändert. Berechnet wird hier eine modifizierte *Levenshtein-Distanz*, die zwischen zwei Segmenten die Editieroptionen ‚delete‘ und ‚add‘, also ‚Abtritt einer Figur‘ und ‚Auftritt einer Figur‘, zulässt. Betrachtet man etwa die erste und die zweite Szene von Goethes *Iphigenie auf Tauris*, dann ist in der ersten Szene nur Iphigenie auf der Bühne; in der zweiten Szene sind es Iphigenie und Arkas. Die *Segment Change Rate* beträgt in diesem Fall 0,5: Es liegt ein *Edit* vor, bei einer Gesamtmenge von zwei involvierten Figuren: $1/2 = 0,5$ (vgl. Abb. 12). Die *Segment Change Rate* zwischen zweiter und dritter Szene von Goethes Drama beträgt 0,66. Es finden zwei *Edits* statt; drei Figuren sind insgesamt involviert: $2/3 = 0,66$ (Abb. 13). Die *Segment Change Rate* ist normalisiert; sie bewegt sich stets zwischen 0 (= keine Änderungen der Figuren auf der Bühne) und 1 (= sämtliche Figuren werden zwischen den Segmenten ausgetauscht).

Aus sämtlichen *Segment Change Rates* eines Dramas wird die *Drama Change Rate* berechnet, für die es zwei Varianten gibt: Die *Drama Change Rate (Mean)* gibt den Mittelwert an, die *Drama Change Rate (SD)* die Standardabweichung. Diese Werte wurden im Folgenden für 331 Dramen aus dem *dIina*-Korpus erhoben, wobei die Reduktion der Korpusgröße sich daraus ergibt, dass die Analysen auf Dramen mit mehr als fünf Segmenten beschränkt wurden. Die entsprechenden Ergebnisse erwiesen sich als historisch weitgehend unauffällig, eindeutige Trends zeichneten sich nicht ab, wie Abb. 14 und 15 zeigen, in denen jedes Kreuz für ein Drama steht (Abszisse: Erscheinungsdatum; Ordinate: Wert für *Drama Change*

⁴⁴Siehe Kathleen M. Carley, „Dynamic Network Analysis“, in: Dies./Ronald L. Breiger/Philippa E. Pattison (Hg.), *Dynamic Social Network Modeling and Analysis. Workshop Summary and Papers*, Washington D.C. 2003, 133–145, <http://www.nap.edu/read/10735/chapter/9> (letzter Aufruf 30.5.2019).

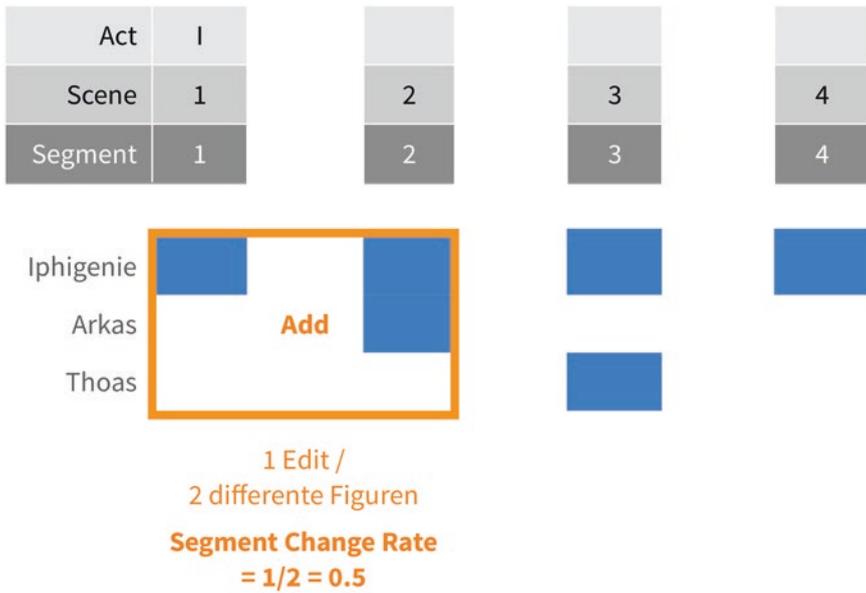


Abb. 12 Berechnung der *Segment Change Rate*, Beispiel 1

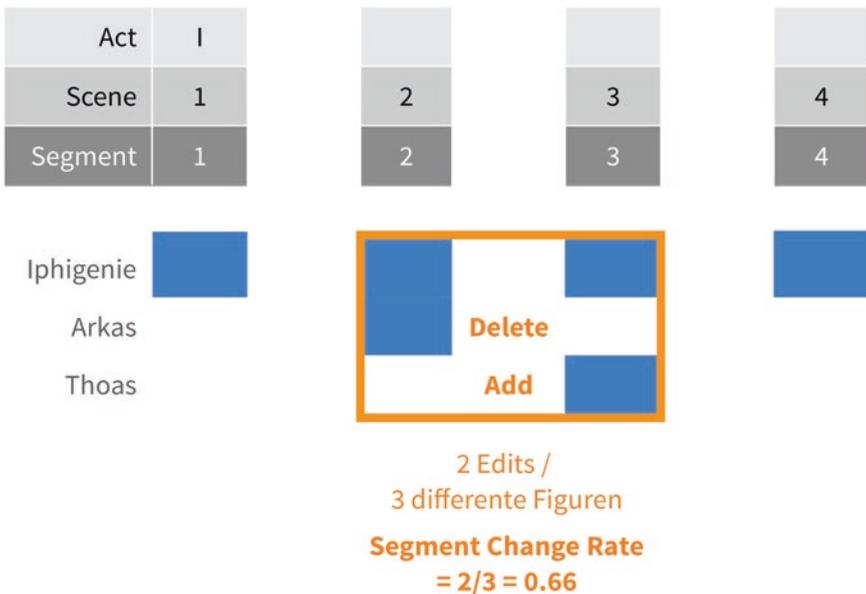


Abb. 13 Berechnung der *Segment Change Rate*, Beispiel 2

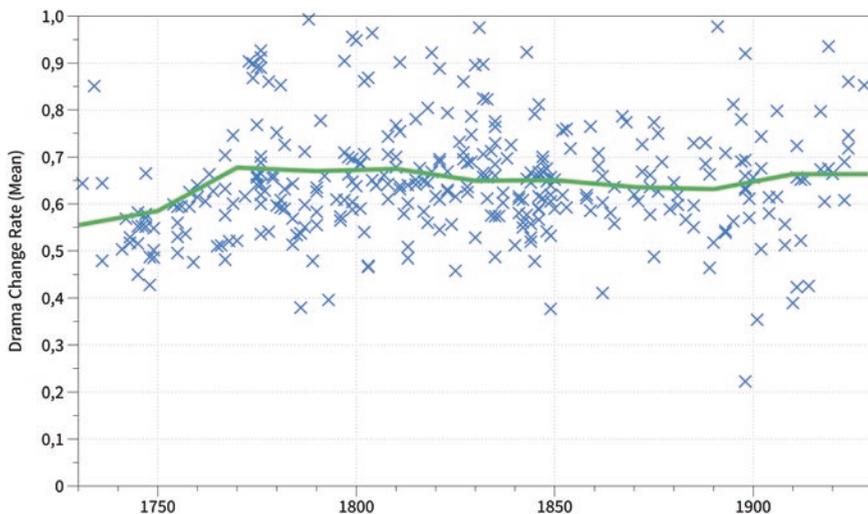


Abb. 14 *Drama Change Rate (Mean)* – Verteilung im *dlina*-Korpus³⁸

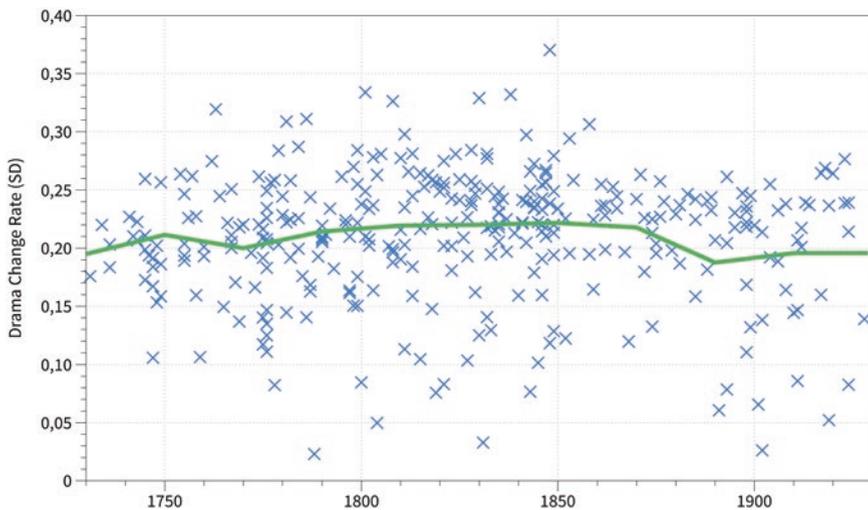


Abb. 15 *Drama Change Rate (SD)* – Verteilung im *dlina*-Korpus⁴⁰

Rate [Mean] bzw. für *Drama Change Rate [SD]*) und die Linien den Mittelwert für alle Dramen eines Zeitraums von jeweils 20 Jahren.

Auffälliger als die historische war die typologische Verteilung, zeigen doch beide Diagramme, dass es zahlreiche Dramen mit extremen Werten gibt. Deren

³⁸ Linie: Mittelwert pro 20 Jahre.

⁴⁰ Linie: Mittelwert pro 20 Jahre.

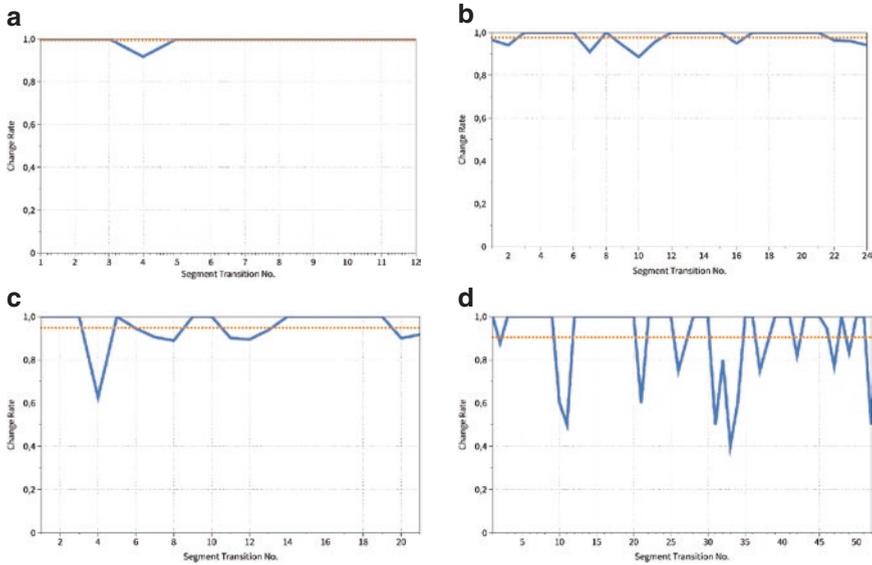


Abb. 16 a. Goethe: *Egmont* (1788) – Mean: 0,99 / SD: 0,02 b. Grabbe: *Napoleon* (1831) – Mean: 0,98 / SD: 0,03 c. Tieck: *Verkehrte Welt* (1800) – Mean: 0,95 / SD: 0,08 d. Goethe: *Götz* (1773) – Mean: 0,90 / SD: 0,17

Betrachtung macht es möglich, eine Typisierung vorzunehmen, bei der jeweils Dramen mit besonders hohen und mit besonders niedrigen Werten als Typen angesetzt werden. Die Eigenart dieser vier Dramen-Typen lässt sich dabei anhand der von uns sog. *Beat Charts* darstellen: Das sind Diagramme, die die Entwicklung der *Segment Change Rate* im Verlauf des Dramas auftragen (durchgehende Linie) und zugleich den Wert für die *Segment Change Rate (Mean)* in einer horizontalen gepunkteten Linie markieren.

Der erste Typ (Abb. 16a, 16b, 16c und 16d), das *hochdynamische Drama*, zeichnet sich durch eine besonders hohe *Drama Change Rate (Mean)* aus; typisch ist hier, dass im Übergang von einem Segment zum nächsten ein großer Teil der Figurenmenge wechselt. Bei Extremtypen wie Goethes *Egmont* (Abb. 16a) wird – mit Ausnahme einer Szene – tatsächlich bei jedem Segmentübergang die gesamte Figurenmenge ausgetauscht.

Der zweite Typ (Abb. 17a, 17b, 28c und 29d), das *niedrigdynamische Drama*, zeichnet sich demgegenüber durch eine sehr niedrige *Drama Change Rate (Mean)* aus; typisch ist hier, dass stets nur ein (kleiner) Teil der Figurenmenge wechselt, mitunter bleibt die Figurenmenge beim Segmentübergang sogar identisch (vgl. die Nullwerte). Die Beispiele, die hier für diesen Typ angeführt sind, weisen dabei eine durchgängige *liasion des scenes* auf, was auch daran liegt, dass dieser Typ vor allem aus Ein- (Abb. 17a und 17b) bzw. Zweiakttern (Abb. 28c und 29d) besteht.

Mit dem dritten Typ (Abb. 18a, 18b, 18c und 18d) richtet sich die Aufmerksamkeit auf die *Drama Change Rate (SD)*, wobei eine hohe Standardabweichung

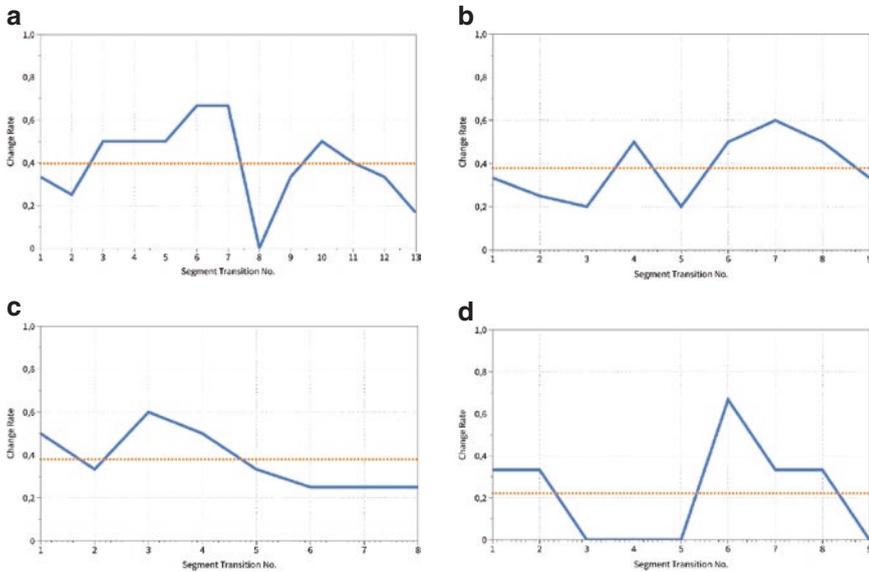


Abb. 17 a. Goethe: *Der Bürgergeneral* (1793) – Mean: 0,40 / SD: 0,18 b. Stephanie: *Der Schauspieldirektor* (1786) – Mean: 0,38 / SD: 0,14 c. Benedix: *Die Hochzeitsreise* (1849) – Mean: 0,38 / SD: 0,13 d. Rilke: *Ohne Gegenwart* (1898) – Mean: 0,22 / SD: 0,22

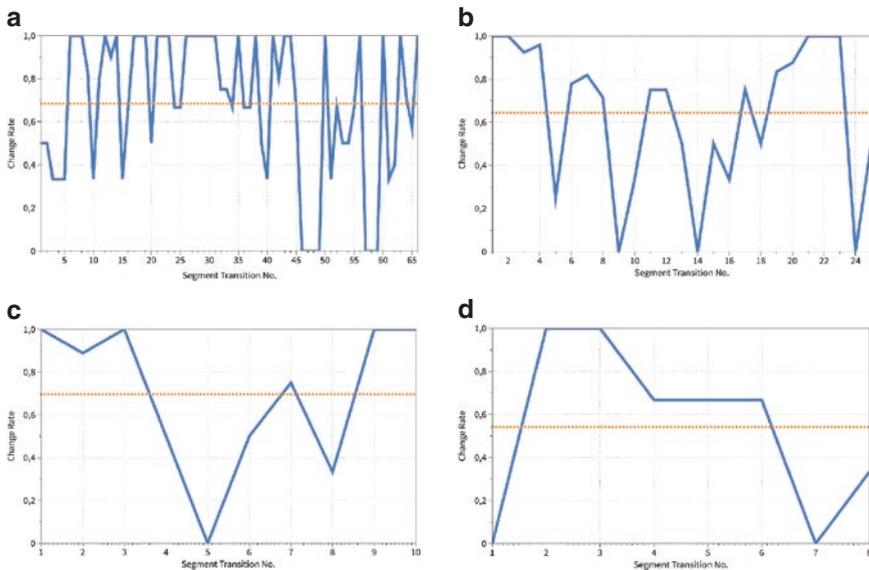


Abb. 18 a. Benkowitz: *Die Jubelfeier der Hölle* (1801) – Mean: 0,68 / SD: 0,33 b. Goethe: *Faust I* (1808) – mean: 0,64 / SD: 0,33 c. Büchner: *Leonce und Lena* (1838) – Mean: 0,70 / SD: 0,33 d. Dulk: *Die Wände* (1848) – Mean: 0,54 / SD: 0,37

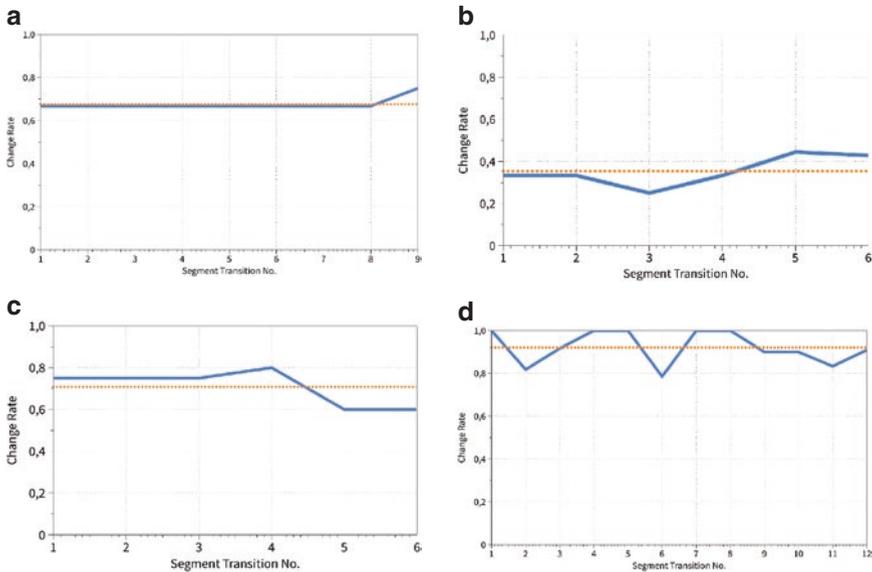


Abb. 19 a. Schnitzler: *Der Reigen* (1902) – Mean: 0,68 / SD: 0,03 b. Thoma: *Die Medaille* (1901) – Mean: 0,35 / SD: 0,07 c. Schnitzler: *Anatol* (1893) – Mean: 0,71 / SD: 0,08 d. Uhland: *Ludwig der Bayer* (1819) – Mean: 0,92 / SD: 0,08

das *starkrhythmische Drama* charakterisiert. Typisch ist hier, dass die Menge der Figuren, die beim Übergang von einem Segment zum folgenden ausgetauscht wird, stark variiert: Mal werden, wie im ersten Beispiel zu sehen (Abb. 18a), sämtliche Figuren ausgetauscht (die durchgezogene Linie berührt den oberen Rand des Diagramms), mal keine (die durchgezogene Linie berührt den unteren Rand), was zu einer starken Rhythmisierung führt.

Beim vierten Typ schließlich (Abb. 19a, 19b, 19c und 19d) – zu dem auch einige der Dramen des ersten Typs, allen voran Goethes *Egmont* zu zählen sind –, dem *schwachrhythmischen Drama*, liegt eine sehr niedrige *Drama Change Rate* (SD) vor. Typisch ist hier, dass stets der tendenziell gleiche (normalisierte) Anteil der Figurenmenge ausgetauscht wird, prototypisch umgesetzt im ‚Reigen‘-Prinzip von Schnitzlers gleichnamigen Drama (Abb. 19a), bei dem – mit Ausnahme des Übergangs zur letzten Szene – stets zwei Figuren auf der Bühne sind, wobei beim Szenenwechsel eine dieser Figuren abtritt und eine neue hinzutritt.

Die vier Extremtypen konturieren einen fünften Typ, den man als *Normaltyp* – mit moderater Dynamik und moderater Rhythmik – bezeichnen kann (Abb. 20a, 20b, 20c und 20d) und der sowohl bei der *Drama Change Rate* (Mean) als auch bei der *Drama Change Rate* (SD) nah an den entsprechenden Mittelwerten für das Gesamtkorpus liegt.⁴⁷

⁴⁷Der Mittelwert für die *Drama Change Rate* (Mean) des Gesamtkorpus liegt bei 0,65. Der Mittelwert für die *Drama Change Rate* (SD) des Gesamtkorpus liegt bei 0,21.

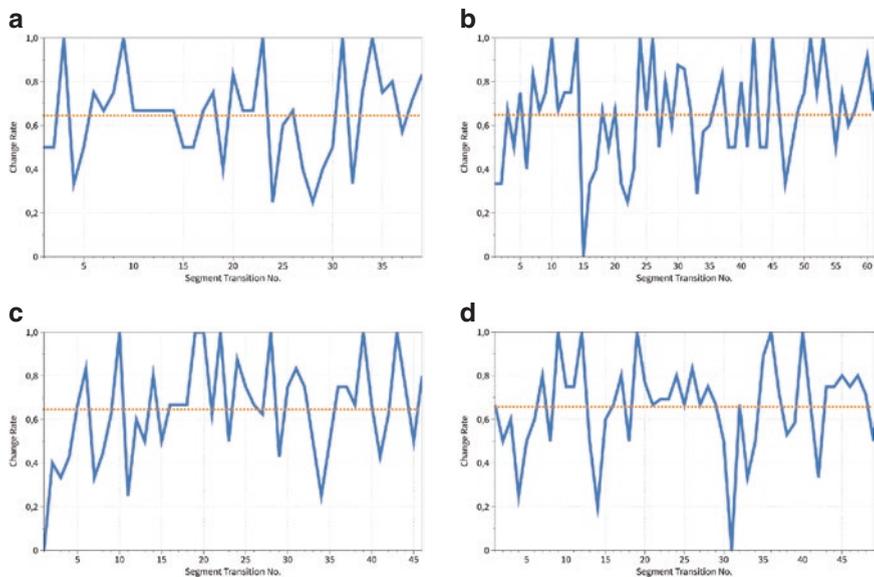


Abb. 20 a. Gottsched: *Die Pietisterei* (1736) – Mean: 0,64 / SD: 0,20 b. Schiller: *Wallensteins Tod* (1799) – Mean: 0,65 / SD: 0,22 c. Ganghofer: *Der Herrgottschnitzer* (1880) – Mean: 0,65 / SD: 0,23 d. Sudermann: *Der Bettler von Syrakus* (1911) – Mean: 0,66 / SD: 0,21

Ein Grund für diese spezifische quantitative Charakteristik des *Normaltyps* zeigt sich, wenn man in den *Beat Chart* die Aktgrenzen einträgt, wie – in Form vertikaler Linien – in Abb. 21 bei Ludwig Ganghofers ‚Volksschauspiel‘ *Der Herrgottschnitzer von Ammergau* geschehen. Hier wird deutlich, dass in diesem Drama Extremausschläge nach oben häufig mit Aktgrenzen zusammenfallen, bei denen es in der klassischen Dramenästhetik zu Ortswechseln oder (geringfügigen) Zeitsprüngen und, damit einhergehend, einem vollständigen Austausch der Figuren kommen darf, wohingegen innerhalb der Akte das Prinzip der *liaison des scenes* dominiert.

Dieser Beobachtung könnte nun systematisch nachgegangen werden: Wie schreiben sich Aktgrenzen in die *Drama Change Rate* ein? Weiterfragen ließe sich dann, ob sich bei klassisch segmentierten Dramen (also, sagen wir: Dramen mit fünf- oder dreiaktiger Struktur und aktinterner Szenenunterteilung) bestimmte Charakteristika von Akten beschreiben lassen, etwa Akte, die typischerweise besonders dynamisch bzw. besonders rhythmisch ausfallen; oder typische Formen der Dynamik bzw. Rhythmik im ersten Akt, die womöglich mit unterschiedlichen Expositionstypen in Verbindung gebracht werden könnten.

Einer solchen Vertiefung der Analysen sei an dieser Stelle jedoch die Reflexion vorgeschaltet. Denn was mit der Frage nach einer netzwerkanalytischen Operationalisierung von Plot begann, mündete – getrieben von der Idee der ‚progressiven Strukturierung‘, dem Maß der *Drama Change Rate* und den *Beat-*

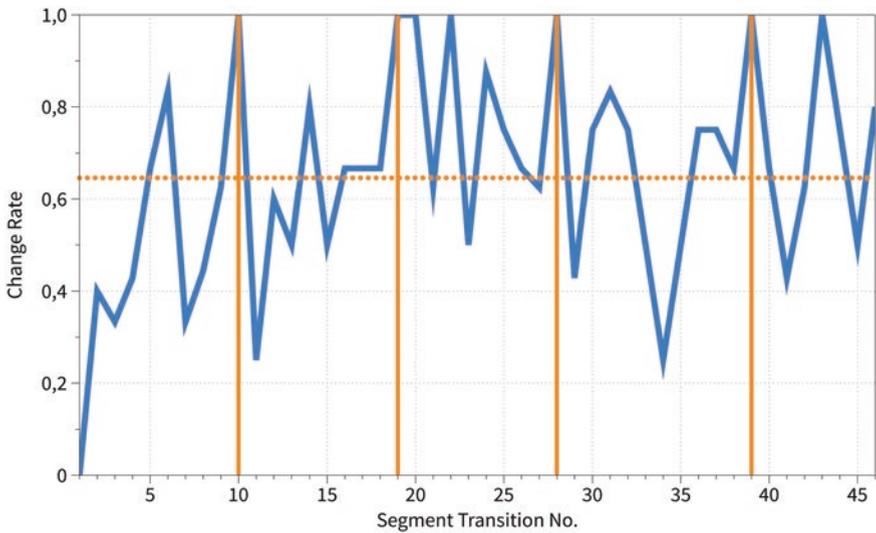


Abb. 21 Ganghofer: *Der Herrgottschnitzer* (1880) – mit Aktgrenzen (vertikale Linien)

Chart-Diagrammen – in einer unvertrauten Beschreibungsweise, von der sich, wie bei der *Small-World*-Studie, in tradierten literaturwissenschaftlichen Begriffen nicht recht sagen lässt, was sie eigentlich beschreibt.

5 Vier Irritationen

Was folgt aus dieser begrifflichen Heimatlosigkeit, in die sich beide Studien sehenden Auges manövriert haben und die sich als eine der „theoretical consequences“ erweist, die aus diesem „encounter of computation and criticism“ folgt? Eine Strategie, sich dieser Frage nach den Konsequenzen zumindest vorerst zu entziehen, bestünde in einer Rhetorik des Aufschubs: *Noch* stehen diese digitalen Textanalysen am Anfang, *noch* bedarf es der systematischen Anwendung der Begriffe und Verfahren auf größere Korpora, *noch* bedarf es der Verfeinerung der Operationalisierungen und der Algorithmen, bedarf es der tieferen Auszeichnung der Texte; und es bedarf *noch* der gründlichen und aufmerksamen Interpretation der Daten, die im Zuge all der Begriffserprobungen und Korpuserkundungen gewonnen wurden und werden. All diese ‚Nochs‘ sind nicht von der Hand zu weisen, jedenfalls mit Blick auf die beiden Studien aus dem *dlna*-Projekt, die hier referiert wurden. Nicht auszuschließen ist dabei, dass einige der Erprobungen und Erkundungen sich in diesem Zustand des ‚Nochs‘ erschöpfen werden – gefangen im (selbstgegrabenem) *Rabbit Hole* der „programming imagination“, wo sie schon bald

dem Vergessen anheimfallen werden, ähnlich einiger Vorgängerarbeiten, in deren Tradition das *dlina*-Projekt verortet werden kann.⁴⁸

Die zu Beginn dieses Essays eingeführte Position von Moretti schert sich allerdings nicht um ein solch mögliches Schicksal, sie besteht auf einer Suspension des Zukünftigen und damit auf einer Ignoranz gegenüber jenen Versprechen auf kommendes Wissen, die in all den ‚Nochs‘ stecken. Stattdessen fordert sie in einer Art metaisierender Verschiebung dazu auf, das, was im Zuge all der zukunftsungewissen Studien geschieht, als konzeptuelle Umbauten am Begriffsgebäude der Literaturwissenschaft zu beobachten und zu reflektieren.

Vier Erfahrungen, vier Irritationen seien hier, daran anschließend, festgehalten.

Dass, *erstens*, Zahlen für die Begriffsarbeit der digitalen Analyse literarischer Texte eine mindestens ebenso große Rolle spielen wie Bedeutungen, ist trivial. Dass statistische Operationen den Bedeutungskern der Begriffe dominieren, ist dennoch bemerkenswert: Die Begriffe der *Digitalen Literaturwissenschaft* sind in Teilen aufgelöst in spezifische Maßzahlen, in *Measures*. Die dramatische *Small World* z. B. definiert sich über zwei netzwerkanalytische Maße, die (einfachen) Signifikanztests unterzogen werden. Im Zuge dessen wird eine tausendfach randomisierte Variante des formalisierten ästhetischen Objekts ‚Drama‘ zum Vergleich herangezogen. Diese statistische Durchdringung der Begriffe, und also der Objekte, hat Konsequenzen: Zum einen führt sie eine relationale Logik in die Begriffsverwendung ein. Begriffe können nur dann auf Objekte angewendet werden, wenn die Daten zu diesen Objekten auf eine definierte, formalisierte Weise in signifikanter Differenz o. Ä. zu den Daten stehen, die zu anderen Objekten erhoben wurden: wenn etwa der *Clustering Coefficient* eines Dramennetzwerks signifikant vom *Clustering Coefficient* der randomisierten Variante dieses Dramennetzwerks abweicht. Diese relationale, statistisch infiltrierte Logik von Begriffsanwendungen, wie sie etwa auch der auf Ähnlichkeitsmaßen basierenden Arbeit mit generischen Begriffen zugrunde liegt⁴⁹ (von den aktuellen *Machine-Learning*-basierten Ansätzen ganz zu schweigen), ist ein Effekt, dessen theoretische Konsequenzen zu reflektieren sind. Zu diesen Konsequenzen gehört, zum anderen, auch, dass die Anwendung von Begriffen stets eine Bezugsgröße erfordert. Im Fall der *Small World*-Studie wie im Fall der *Change Rate*-Typologisierung ist dies das Korpus, dessen Teil das einzelne Objekt ist. Anders gesagt: Der Begriff der *dramatischen Small World* kann auf Goethes *Götz* nur in Bezug auf das *dlina*-Korpus angewendet werden, zu dem dieser Text hinsichtlich der entscheidenden Werte in signifikanter Abweichung steht. Und ein *Change Rate*-Normaltypus wie Ganghofers *Herrgottschnitzer* ist ‚normal‘ zunächst nur innerhalb jenes Korpus, zu dem es in Bezug gesetzt wurde. Damit

⁴⁸Verwiesen sei hier exemplarisch auf Solomon Marcus, *Mathematische Poetik*, Frankfurt a. M. 1973; Waltraud Reichert, *Informationsästhetische Untersuchungen an Dramen*, Stuttgart 1965.

⁴⁹Exemplarisch Christof Schöch, „Zeta für die kontrastive Analyse literarischer Texte“, in: Toni Bernhart/Marcus Willand/Sandra Richter/Andrea Albrecht (Hg.), *Quantitative Verfahren in den Literatur- und Geisteswissenschaften. Systematische und historische Perspektiven*, Berlin/Boston 2018, 77–94.

aber erweist sich die statistische Durchdringung der Begriffsarbeit in der *Digitalen Literaturwissenschaft* wesentlich geprägt durch eine numerische, an quantifizierten Ähnlichkeits-, Abstands- oder Differenzmaßen orientierte Variante der Logik aus Norm und Abweichung.⁵⁰ Und als zutiefst durchdrungen von jenen literaturwissenschaftlichen Fragen der Korpusbildung, die in jüngerer Zeit unter dem Titel ‚Wertung und Kanon‘ diskutiert werden.

Im Grunde bereits impliziert im ersten Punkt ist, *zweitens*, eine radikale ‚Anästhetisierung‘ der Objekte – also der literarischen Kunstwerke – im Zuge der Begriffsarbeit. Selbst dort, wo es, wie in den *dlna*-Studien, um Aspekte der Komposition geht, verschwindet das ästhetische Objekt, weil es im *Distant Reading* der Erfahrbarkeit als literarisches Werk konstitutiv entzogen ist. Eine ähnliche Anästhetisierung lässt sich für den Stil-Begriff beobachten.⁵¹ Dabei bedeutet die Anästhetisierung des Objekts keineswegs, dass dieses nicht in seinem spezifischen Artefakt-Charakter betrachtet wird. Wenn etwa in der *Small World*-Studie die Spezifik der dramatischen Netzwerke in Abhebung zu deren randomisierten Varianten untersucht wird, dann zielt eine solche Analyse gerade darauf, die Nicht-Zufälligkeit und also die Intentionalität der Formung zu erfassen, ihr Gestaltetsein.

Der Anästhetisierung der Objekte korrespondiert, *drittens*, eine Aisthetisierung der Begriffsanwendung, ihre Diagrammatisierung.⁵² Dieser Aspekt der digitalen Analyse und ihrer Visualität wird derzeit u. a. von Johanna Drucker kritisch reflektiert.⁵³ Auch Bildlichkeit ist für die Begriffsarbeit deshalb von besonderer Bedeutung, weil das Anschaulichwerden der Daten in der Visualisierung stets eine Interpretation ist, in der Objekt und Begriff auf spezifische Weise zusammenkommen. Die oben abgebildeten *Beat Charts* etwa fungieren als Mittel, durch das die Maßzahlen der *Segment Change Rate* und der *Drama Change Rate* mit den typologischen Begriffen des *hochdynamischen*, *niedrigdynamischen* etc. *Dramas* visuell amalgamiert werden: Die *Beat Charts* sind gewissermaßen das Medium, das den gemessenen Daten Sinn und ‚Sinnlichkeit‘ geben soll – das Begriffe sinnlich und Anschauungen (Daten) verständlich macht. Das aber ist, darauf hat Drucker mit Nachdruck hingewiesen, alles andere als ein konventionalisierter Modus der Begriffsverwendung in den Geisteswissenschaften, wobei keineswegs geklärt ist, inwieweit sich die Konventionalisierung der diagrammatischen Begriffsarbeit in der *Digitalen Literaturwissenschaft* einfach in der Übernahme jener Konventionen erschöpfen kann, die etwa die Natur- oder Sozialwissenschaften ausgebildet haben.

⁵⁰ Siehe Harald Fricke, *Norm und Abweichung. Eine Philosophie der Literatur*, München 1981.

⁵¹ Siehe grundlegend J. Berenike Herrmann/Karin Dalen-Oskam/Christof Schöch, „Revisiting Style, a Key Concept in Literary Studies“, in: *Journal of Literary Theory* 9.1 (2015), 25–52.

⁵² Verstanden in einem weiten Sinne, vgl. etwa Sybille Krämer, *Figuration, Anschauung, Erkenntnis. Grundlinien einer Diagrammatologie*, Berlin 2016; Birgit Schneider/Christoph Ernst/Jan Wöpking (Hg.), *Diagrammatik-Reader*, Berlin/Boston 2016.

⁵³ Johanna Drucker, *Graphesis. Visual Forms of Knowledge Production*, Cambridge, MA, 2014.

Was schließlich, *viertens*, irritiert wie fasziniert, sind Modus, Duktus und Gestus der Begriffsarbeit, wie sie in Studien aus dem Feld der digital-empirischen Literaturanalyse vorherrschen, und das unabhängig von der Tatsache, dass diese Begriffsarbeit in der Regel statistisch, anästhetisch und diagrammatisch ist. Denn es gibt auch eine Art *Rhetorik der digital-empirischen Literaturanalyse*. Dies betrifft auf der einen Seite jene ‚Paper‘ der *Digitalen Literaturwissenschaft*, die, in streng empirischer Manier, aus argumentativer, stilistischer und layouterischer Sicht kaum zu unterscheiden sind von den Veröffentlichungen der Science-Disziplinen. Es betrifft auf der anderen Seite auch jene *Pamphlets*, Blogposts, Github-Repositories, Korpus-Server usw., die dem Versuchsweisen, dem Unfertigen, dem Offenen verpflichtet sind. Gemeinsam ist beiden Publikationsformen, so scheint mir, eine Verschiebung vom Ergebnis zum Prozessualen, auch zum Performativen. Nicht der Abschluss der wissenschaftlichen Untersuchung, sondern ihr Aufschub: ihr Suchen und Versuchen prägt die Veröffentlichungen der *Digitalen Literaturwissenschaft*. Als hätte man *noch* nicht so recht begriffen, wie sich Argumentationen zu dieser Art von Daten entwickeln lassen, in welche Narrationen diese literarischen Daten eingebunden, mit welchen Darstellungsformen sie in welche Communities kommuniziert werden sollen. Als wäre *noch* unsicher, worauf diese interdisziplinäre, kollaborative, aber im Ansatz philologische Arbeit an Methoden und Begriffen eigentlich hinauslaufen könnte.

Man mag die statistische, anästhetische, diagrammatische, rhetorische Alterität der *Digitalen Literaturwissenschaft* aus guten Gründen zurückweisen. Die Literaturgeschichte geändert haben wir bisher gewiss nicht. Was sich im *Rabbit Hole* der „programming imagination“ ereignete, mag eine Verwirrung – vielleicht eine grundsätzliche, vielleicht eine vorübergehende – der Art, des Stils, der Weise sein, wie Literaturwissenschaft betrieben wird. Um Ausschließlichkeit ginge es dabei nicht. Niemand hat die Absicht, das hermeneutische, philologische, kritische o. ä. Paradigma der Literaturwissenschaft abzulösen, das ästhetische, diskursive, kommunikative o. ä. Objekt zum Verschwinden zu bringen.

Eher geht es darum, die ohnehin bestehende Multidimensionalität der Begriffe auszuweiten, und damit auch die Interpretationswürdigkeit der Objekte, mit denen sich die Literaturwissenschaft befasst, weiter – nun quantitativ – auszureizen. Was *das* ‚Drama‘ ist, hat sich für uns im Verlauf der Studien jedenfalls keineswegs reduziert, sondern erweitert. Es ist uns zuletzt unmöglich geworden, ein Drama nicht auch als eine Netzwerkstruktur zu begreifen. Ob das eine Bereicherung ist, mögen jene beurteilen, die vor dem *Rabbit Hole* sitzen.

Ergänzung 2021

Die Studien der *dlna*-Arbeitsgruppe, deren Resümee der hier publizierte Aufsatz von 2017 in gewisser Hinsicht bereits zog, waren der Idee des Rapid Prototyping verpflichtet. Korpus und Software wurden entlang einer Reihe von experimentellen Studien er-, über- und ausgearbeitet, mit denen wir ausloten wollten, welche Beobachtungen, Fragen und Forschungsszenarien sich aus der Einführung von Dramen mit den Erkenntnispraktiken der Netzwerkanalyse ergeben – wobei wir uns dynamisch, auch sprunghaft der wissenschaftlichen Neu-

gier überließen. Wie sich das Dramen-Korpus und die v. a. von Christoph Kittel verantwortete Software *dramavis* im Verlauf der Studien aus den Jahren 2015 bis 2018 entwickelten, war also weniger von einem übergreifenden Plan oder einer langfristigen Strategie bestimmt, sondern vielmehr von der Attraktion des jeweiligen Erkenntnisinteresses getrieben, das sich ganz wesentlich auch aus den Irritationspotenzialen des epistemischen Objektes ‚Dramennetzwerk‘ speiste.⁵⁴

Die Irritationspotenziale normalisierten sich, das Erkenntnisinteresse weitete sich. Schon die folgende *dlina*-Studie zur quantitativen Figurentypologie (s. Fischer/Trilcke/Kittel u. a. 2018)⁵⁵ argumentierte sowohl mit Netzwerkmaßen *als auch* mit wortfrequenzbasierten Maßzahlen; eine Studie zu den Bühnenanweisungen (s. Trilcke/Kittel/Reiter/Maximova/Fischer 2020) operierte schließlich überhaupt nicht mehr mit den Verfahren der Netzwerkanalyse, sondern griff auf Methoden und Tools aus dem *Natural Language Processing* zurück, u. a. auf das im QuaDrama-Projekt erarbeitete Tool *DramaNLP* (Reiter/Pagel 2020).

Mit der Pluralisierung der computationellen Methoden unserer Dramenforschung, die sich in diesen Studien abzeichnete, konnte unser Korpus aus präzise edierten Strukturdaten (in Gestalt des *dlina*-‚Zwischenformats‘) nicht mehr mithalten. Was man als die ‚Korpuskrise‘ der *dlina*-Arbeitsgruppe bezeichnen könnte, war dabei eine der Urszenen für die Aufnahme der Arbeiten am *Drama Corpora Project* (*DraCor*),⁵⁶ das auf dem Konzept der ‚Programmable Corpora‘ beruht (s. Fischer/Börner/Göbel u. a. 2019). Mit *DraCor* begann letztlich alles noch einmal neu.

Dabei sind zahlreiche Forschungen und Entwicklungen der *dlina*-Studien in die *DraCor*-Plattform eingegangen, bis hin zu den Metriken, die die in *DraCor* eingebauten Dienste berechnen. Zugleich aber setzt *DraCor* eben noch einmal neu an, tritt auch einen Schritt zurück und blickt in Richtung einer Infrastruktur, die auf die Ermöglichung breiterer, multidimensionaler (Dramen-)Forschung zielt – und sich von Beginn an einer ambitionierten Idee von reproduzierbarer Forschung verpflichtet, die nicht zuletzt auf der Bereitstellung einer spezifischen API gründet.

⁵⁴Dieses Irritationspotenzial ist wesentlich ein Effekt der strukturellen Abstraktion und der aus dieser resultierenden Möglichkeiten des analytischen Umgangs mit den epistemischen Objekten, zu denen nicht zuletzt die in der Netzwerkanalyse typische randomisierte Generierung von Netzwerken gehört, wie wir sie in der Small World-Studie eingesetzt haben. In seiner Studie „Simulating Dramatic Networks“ hat Franco Moretti (2020) mit m.W.n. bisher größter Konsequenz versucht, diesen generativen Ansatz der Netzwerkanalyse für literaturwissenschaftliche Fragen nutzbar zu machen.

⁵⁵Vgl. mit ähnlichem Ansatz auch Benjamin Krautter/Janis Pagel/Nils Reiter u. a., *Titelhelden und Protagonisten – Interpretierbare Figurenklassifikation in deutschsprachigen Dramen* (= LitLab Pamphlets 7). o. O. 2018, https://www.digitalhumanitiescooperation.de/wp-content/uploads/2018/12/p07_krautter_et_al.pdf (letzter Aufruf 31.08.2021); vgl. zudem, etwas anders ausgerichtet, auch Marcus Willand/Benjamin Krautter/Janis Pagel u. a., „Passive Präsenz tragischer Hauptfiguren im Drama“, in: *DHd2020. Abstracts*, o. O. 2020. <https://doi.org/10.5281/zenodo.4621812>.

⁵⁶<https://dracor.org/>

Mittlerweile ist auch *DraCor* eingebettet in einen größeren Kontext. Seit März 2021 ist die in *DraCor* prototypisch realisierte Idee der ‚Programmable Corpora‘ eines der konzeptionellen Zentren, die in dem von der Europäischen Kommission geförderten Horizon 2020-Projekt „CLS INFRA. Computational Literary Studies Infrastructure“⁵⁷ erforscht und entwickelt werden.

Literatur⁵⁸

- Albert, Réka/Barabási, Albert-László, „Statistical Mechanics of Complex Networks“, in: *Reviews of Modern Physics* 74 (2002), 47–97.
- Algee-Hewitt, Mark, „Distributed Character: Quantitative Models of the English Stage, 1500–1920“, in: *Digital Humanities 2017. Conference Abstracts*, Montréal 2017, 119–121, <https://dh2017.adho.org/abstracts/103/103.pdf>.
- Carley, Kathleen M., „Dynamic Network Analysis“, in: Dies./Ronald L. Breiger/Philippa E. Pattison (Hg.), *Dynamic Social Network Modeling and Analysis. Workshop Summary and Papers*, Washington D.C. 2003, 133–145, <http://www.nap.edu/read/10735/chapter/9>.
- Dannenberg, Hilary P., „Plot“, in: David Herman/Manfred Jahn/Marie-Laure Ryan (Hg.), *The Routledge Encyclopedia of Narrative Theory*, London 2005, 435–439.
- Drucker, Johanna, *Graphesis. Visual Forms of Knowledge Production*, Cambridge, MA, 2014.
- Fischer, Frank/Trilcke, Peer/Kittel, Christopher u. a., „To catch a protagonist: Quantitative dominance relations in german-language drama (1730–1930)“, in: *Digital Humanities 2018. Conference Abstracts, El Colegio de México & Universidad Nacional Autónoma de México & Red de Humanidades Digitales 2018*, 193–201.
- Fischer, Frank/Börner, Ingo/Göbel, Mathias u. a., „Programmable Corpora. Die digitale Literaturwissenschaft zwischen Forschung und Infrastruktur am Beispiel von DraCor“, in: *DHd 2019. Digital Humanities. multimedial & multimodal. Konferenzabstracts*, Mainz/Frankfurt a. M., 194–197, DOI <https://doi.org/10.5281/zenodo.2596094>.
- Fischer, Frank/Göbel, Mathias/Kampkaspar, Dario u. a., „Network Dynamics, Plot Analysis. Approaching the Progressive Structuration of Literary Texts“, in: *Digital Humanities 2017. Conference Abstracts*, Montréal 2017, 437–441, <https://dh2017.adho.org/abstracts/071/071.pdf>.
- Fischer, Frank/Trilcke, Peer, *Introducing dlina Corpus 15.07 (Codename: Sydney). 20. Juni 2015*, <https://dlina.github.io/Introducing-DLINA-Corpus-15-07-Codename-Sydney/>.
- Fischer, Frank/Trilcke, Peer/Kittel, Christopher u. a., „To Catch a Protagonist. Quantitative Dominance Relations in German-Language Drama (1730–1930)“, in: *Digital Humanities 2018. Book of Abstracts* (2018), 193–201, https://dh2018.adho.org/wp-content/uploads/2018/06/dh2018_abstracts.pdf.
- Fischer, Frank/Börner, Ingo/Göbel, Mathias u. a., „Programmable Corpora. Introducing DraCor, an Infrastructure for the Research on European Drama“, in: *Digital Humanities 2019. Book of Abstracts* (2019). DOI: <https://doi.org/10.5281/zenodo.4284002>.
- Freeman, Linton C., *The Development of Social Network Analysis. A Study in the Sociology of Science*, North Charleston 2004.
- Fricke, Harald, *Norm und Abweichung. Eine Philosophie der Literatur*, München 1981.
- Herrmann, J. Berenike/Dalen-Oskam, Karin/Schöch, Christof, „Revisiting Style, a Key Concept in Literary Studies“, in: *Journal of Literary Theory* 9.1 (2015), 25–52.

⁵⁷ <https://clsinfra.io/>, Fördernummer: 101004984 (<https://cordis.europa.eu/project/id/101004984>, letzter Aufruf 26.10.2021).

⁵⁸ Sämtliche digitalen Referenzen wurden letztmalig am 30.5.2019 beziehungsweise am 31.08.2021 eingesehen.

- Hettinger, Lena/Becker Martin/Reger Isabella u. a., „Genre classification on German novels“, in: *Proceedings of the 12th International Workshop on Text-based Information Retrieval*, Valencia 2015, <http://www.uni-weimar.de/medien/webis/events/tir-15/tir15-papers-final/Hettinger2015-tir-paper.pdf>.
- Huyssen, Andreas, *Das Drama des Sturm und Drang. Kommentar zu einer Epoche*, München 1980.
- Jannidis, Fotis, „Digitale Literaturwissenschaft“, in: *Digitalität. Theorien und Praktiken des Digitalen in den Geisteswissenschaften* (4. Juli 2016), <https://digigeist.hypothesen.org/114>.
- Kampkaspar, Dario/Fischer, Frank/Trilcke, Peer, „Introducing Our ‚Zwischenformat‘“, in: *Digital Humanities and Literary Studies* (21. Juni 2015), <https://dlina.github.io/Introducing-Our-Zwischenformat/>.
- Kampkaspar, Dario/Trilcke, Peer, „Editing Rules“, in: *Digital Humanities and Literary Studies* (22. Juni 2015), <https://dlina.github.io/Editing-Rules/>.
- Katsma, Holst, „Loudness in the Novell“, in: *Literary Lab Pamphlet 7* (September 2014), <https://litlab.stanford.edu/LiteraryLabPamphlet7.pdf>.
- Kittel, Christopher/Fischer, Frank, „dramavis (v0.4)“, in: *GitHub Repository* (September 2017), <https://github.com/lehkost/dramavis>.
- Klotz, Volker, *Geschlossene und offene Form im Drama*, München ³1968.
- Krämer, Sybille, *Figuration, Anschauung, Erkenntnis. Grundlinien einer Diagrammatologie*, Berlin 2016.
- Krautter, Benjamin/Pagel, Janis/Reiter, Nils u. a., *Titelhelden und Protagonisten – Interpretierbare Figurenklassifikation in deutschsprachigen Dramen* (= LitLab Pamphlets 7). o. O. 2018, https://www.digitalhumanitiescooperation.de/wp-content/uploads/2018/12/p07_krautter_et_al.pdf.
- Kukkonen, Karin, „Plot“, in: Peter Hühn/Jan Christoph Meister/John Pier/Wolf Schmid (Hg.), *the living handbook of narratology* (24. März 2014), <http://www.lhn.uni-hamburg.de/article/plot>.
- Marche, Stephen, „Literatur Is not Data. Against Digital Humanities“, in: *Los Angeles Review of Books* (28. Oktober 2012), <https://lareviewofbooks.org/article/literature-is-not-data-against-digital-humanities/>.
- Marcus, Solomon, *Mathematische Poetik*, Frankfurt a. M. 1973.
- Moretti, Franco, „Operationalizing“. Or, the function of measurement in modern literary theory“, in: *Literary Lab Pamphlet 6* (Dezember 2013), <https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf>.
- Moretti, Franco, „Conjectures on World Literature“, in: Ders., *Distant Reading*, London/New York 2013, 43–62.
- Moretti, Franco, „Literature, Measured“, in: *Literary Lab Pamphlet 12* (April 2016), <https://litlab.stanford.edu/LiteraryLabPamphlet12.pdf>.
- Moretti, Franco, „Network Theory, Plot Analysis“, in: *Literary Lab Pamphlet 2* (Mai 2011), <http://litlab.stanford.edu/LiteraryLabPamphlet2.pdf>.
- Moretti, Franco, „Simulating Dramatic Networks. Morphology, History, Literary Study“, in: *Journal of World Literature* (2020), DOI: <https://doi.org/10.1163/24056480-20201001>.
- Pfister, Manfred, *Das Drama. Theorie und Analyse*, München ⁹1997.
- Reichert, Waltraud, *Informationsästhetische Untersuchungen an Dramen*, Stuttgart 1965.
- Reiter, Nils/Pagel, Jannis: *DramaNLP*. v2.1.4. (2020). DOI: <https://doi.org/10.5281/zenodo.3897598>.
- Schneider, Birgit/Ernst, Christoph/Wöpking, Jan (Hg.), *Diagrammatik-Reader*, Berlin/Boston 2016.
- Schöch, Christof, „Zeta für die kontrastive Analyse literarischer Texte“, in: Toni Bernhart/Marcus Willand/Sandra Richter/Andrea Albrecht (Hg.), *Quantitative Verfahren in den Literatur- und Geisteswissenschaften. Systematische und historische Perspektiven*, Berlin/Boston 2018, 77–94.
- Stiller James/Nettle Daniel/Dunbar Robin I. M., „The Small World of Shakespeare’s Plays“, in: *Human Nature* 14 (2003), 397–408.
- Stiller, James/Hudson, Matthew, „Weak Links and Scene Cliques Within the Small World of Shakespeare“, in: *Journal of Cultural and Evolutionary Psychology* 3 (2005), 57–73.
- Trilcke, Peer, „Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft“, in: Philip Ajouri/Katja Mellmann/Christoph Rauen (Hg.), *Empirie in der Literaturwissenschaft*, Münster 2013, 201–247.

- Trilcke, Peer/Fischer, Frank, „Literaturwissenschaft als Hackathon. Zur Praxeologie der digitalen Literaturwissenschaft und zu ihren epistemischen Dingen“, in: Martin Huber/Sybille Krämer (Hg.), *Wie Digitalität die Geisteswissenschaften verändert. Neue Forschungsgegenstände und Methoden*, Sonderband der Zeitschrift für digitale Geisteswissenschaft 3 (27. Juni 2018), DOI https://doi.org/10.17175/sb003_003.
- Trilcke, Peer/Fischer, Frank/Göbel, Mathias u. a., „Dramen als ‚Small Worlds‘? Netzwerkdaten zur Geschichte und Typologie deutschsprachiger Dramen 1730–1930“, in: *DHd 2016. Modellierung – Vernetzung – Visualisierung. Konferenzabstracts*, Leipzig 2016, 255–258, <http://dhd2016.de/boa.pdf>.
- Trilcke, Peer/Fischer, Frank/Göbel, Mathias u. a., „Theatre Plays as ‚Small Worlds‘? Network Data on the History and Typology of German Drama, 1730–1930“, in: *Digital Humanities 2016. Conference Abstracts*, Kraków 2016, 385–387, <http://dh2016.adho.org/abstracts/360>.
- Trilcke, Peer/Fischer, Frank/Göbel, Mathias, „Netzwerkdyamik, Plotanalyse. Zur Visualisierung und Berechnung der ‚progressiven Strukturierung‘ literarischer Texte“, in: *DHd 2017. Digitale Nachhaltigkeit. Konferenzabstracts*, Bern 2017, 175–180, http://www.dhd2017.ch/wp-content/uploads/2017/03/Abstractband_def3_M%C3%A4rz.pdf.
- Trilcke, Peer/Kittel, Christopher/Reiter, Nils u. a., „Opening the Stage. A Quantitative Look at Stage Directions in German Drama“, in: *Digital Humanities 2020. Conference Abstracts* (2020), https://dh2020.adho.org/wp-content/uploads/2020/07/337_OpeningtheStageAQuantitativeLookatStageDirectionsinGermanDrama.html.
- Watts, Duncan J./Strogatz, Steven H., „Collective Dynamics of ‚Small World‘ Networks“, in: *Nature* 393 (1998), 440–442.
- Willand, Marcus/Krautter, Benjamin/Pagel, Janis u. a., „Passive Präsenz tragischer Hauptfiguren im Drama“, in: *DHd2020. Abstracts*, o. O. 2020. DOI: <https://doi.org/10.5281/zenodo.4621812>.

Online-Ressourcen

- (Digital) Literary Network Analysis, <https://dlina.github.io/>.
 DraCor, <https://dracor.org>.
 GitHub Repository, dlina, <https://github.com/dlina>.
 Shiny DraCor, <https://shiny.dracor.org/>.
 TextGrid Repository, <https://textgrid.de/>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.



***Burrows's Delta* und Z-Score-Differenz im Netzwerkvergleich. Analysen zum *Deutschen Novellenschatz* von Paul Heyse und Hermann Kurz (1871–1876)**

Thomas Weitin

1 Methodische Vorüberlegungen

In unserer wissenschaftlichen Arbeit schätzen wir die Theorie bloß als Arbeitshypothese, mittels derer die Fakten entdeckt werden und einen Sinn bekommen: in ihrer Gesetzmäßigkeit sowie als Material der Forschung. [...] Wir stellen konkrete Grundsätze auf und halten uns daran, sofern sie vom Material verifiziert werden.

Boris Ejchenbaum¹

Zu den Innovationen *Digitaler Literaturwissenschaft* gehört die Entwicklung methodenkritischer Perspektiven für das eigene Fach, die grundsätzlich die Frage betreffen, wie wissenschaftlich gedacht und analysiert werden soll. Mehr als die anderen hermeneutischen Textwissenschaften ist die Literaturwissenschaft in den Bereichen, die der *Cultural Turn* geprägt hat, von einem Theorienpluralismus bestimmt worden, in dem sich die Ansätze nur selten wechselseitig erhellen konnten. Eine Ursache dafür ist die unterschwellige Ähnlichkeit im methodischen Vorgehen, die von den theoretischen Eigenheiten weitgehend verdeckt worden ist. Ohne viel Beachtung und selbstverständlich mit nicht wenigen Ausnahmen konnte ein verifikationsorientierter Denkstil dominant werden, der darauf zielte, die jeweilige These oder Theorie in der Textinterpretation zu belegen. Dagegen ist grundsätzlich nichts einzuwenden, Stellenlektüre gehört zum kritischen Inventar hermeneutischer Techniken. Wichtig ist aber, sich der Tatsache bewusst zu sein, dass es sich dabei um ein bestimmtes methodisches Vorgehen handelt, dessen

¹Boris Ejchenbaum, „Zur Theorie der formalen Methode“, in: Ders., *Aufsätze zur Theorie und Geschichte der Literatur*, Frankfurt a. M. 1965, 7–52, hier: 8.

T. Weitin (✉)
Technische Universität Darmstadt, Darmstadt, Deutschland
E-Mail: Thomas.Weitin@tu-darmstadt.de

Nutzen wie bei jedem anderen Verfahren einen Preis hat. Wer verifikationsorientiert denkt, wird das, was die Theorie stützt, immer stärker würdigen als das, was nicht dazu passt.

Zweifellos ist ein solcher *Confirmation Bias* nicht auf die Geisteswissenschaften beschränkt. Die Entwicklung der Literaturwissenschaft zur digitalen Fachwissenschaft macht es aber in jedem Fall notwendig, auch alternative Ansätze zu entwickeln. Digitale Textanalyse bedarf einer empirischen Perspektive und muss statistische Zusammenhänge berücksichtigen, was wiederum einen anderen Umgang mit eigenen Theorien verlangt. Als ‚wissenschaftlich‘ wird eine These demnach angesehen, wenn Daten denkbar sind, die sie als falsch erweisen. Oder anders gewendet: mit Thesen, die nicht an der Realität des untersuchten Materials scheitern können, lässt sich kein kritischer Diskurs über etwas führen.² Das Risiko, falsch zu liegen, und die Möglichkeit von Kritik gehören unmittelbar zusammen. Weil das so ist, sollten digitale Analysen auch nicht rein explorativ verfahren. In fast jeder Studie sind explorative Phasen und Zwischenschritte notwendig, um einen neuen Gedanken, ein Modell zu testen oder die Parameter eines Tools zu optimieren. Wer aber im Ganzen oder generell so verfährt, reproduziert unweigerlich das Problem, das allen explorativen Untersuchungen anhaftet: Sie werden so lange durchgeführt, bis man mit dem Ergebnis zufrieden ist. Auch das ist keine günstige Voraussetzung für einen kritischen Diskurs in der Wissenschaft. Geeigneter dafür erscheint ein hypothesengeleitetes Vorgehen, das sich empirisch auf Theorien stützt, die dann scheiternd oder sich bewährend gleichermaßen von Interesse sind.

Unsere hypothetischen Überlegungen zum *Deutschen Novellenschatz* der Herausgeber Paul Heyse und Hermann Kurz haben ihren Ausgangspunkt beim Status und der Selbstverortung dieser historischen Sammlung innerhalb der literarischen Kommunikation ihrer Zeit (1871–1876) genommen.³ Sie umfasst 86 Texte, die nach in der Einleitung niedergelegten Prinzipien („Falkentheorie“) von den Herausgebern unter eindeutiger Federführung Heyses als Novellen klassifiziert worden sind. Diese Theorie hat uns bislang hauptsächlich als eine realistische Gattungspoetik interessiert, die auf zwei grundlegende Fragen im Zeitalter literarischer Massenproduktion zu antworten versucht: Wie schreibt man im Bewusstsein massenhafter Ähnlichkeit auf dem literarischen Markt, wie kann man sich überhaupt noch unterscheiden? Und wie schreibt man unter diesen Bedingungen Literaturgeschichte?⁴ Die Idee hinter der Falkentheorie ist

²Vgl. Norbert Groeben, „Was kann/soll ‚Empirisierung in der Literaturwissenschaft‘ heißen?“, in: Philip Ajouri/Katja Mellmann/Christoph Rauen (Hg.), *Empirie in der Literaturwissenschaft*, Münster 2013, 47–74, hier: 53 f.

³Ich arbeite daran nicht allein, sondern im Rahmen einer von der *VolkswagenStiftung* im Programm *Mixed Methods in the Humanities* geförderten Forschergruppe gemeinsam mit Ulrik Brandes (ETH Zürich). Unser Projekt trägt die Förderungsnummer 91 920.

⁴Vgl. Thomas Weitin, „Selektion und Distinktion. Paul Heyses und Hermann Kurz’ *Deutscher Novellenschatz* als Archiv, Literaturgeschichte und Korpus“, in: Daniela Gretz/Nicolas Pethes

ein einfacher Test. Die „einfache Form“ der Novelle soll so genutzt werden, dass „das Spezifische“ eines Sujets herausgearbeitet wird, „das diese Geschichte von tausend anderen unterscheidet“.⁵ Dabei entwickelt sich im Bewusstsein der notwendigen Distinktion auf dem literarischen Massenmarkt eine zeitgemäße Vorstellung von Originalität, die sich differenziell als Distanzmaß des einzelnen Textes zum Gesamtkorpus der Gattung bestimmt. Diese handreichungsartige Novellentheorie steht in systematischem Zusammenhang mit den Prinzipien der Textauswahl für den *Novellenschatz*. So, wie die Schriftstellerinnen und Schriftsteller angehalten werden, ihren jeweiligen Text schon in dem Bewusstsein sehr vieler ähnlicher Texte zu verfassen, sehen sich die Herausgeber auf der Suche nach geeigneten Novellen mit überwältigenden Textmassen konfrontiert. Das ursprüngliche Konzept einer chronologischen Literaturgeschichte der deutschsprachigen Novellistik scheitert daran. An seine Stelle tritt ein anderer Ansatz, der auf der Basis der formalen Gemeinsamkeiten, die die Novellengattung verbürgt, Stilveränderungen zwischen den Epochen – vor allem zwischen Romantik und Realismus – beobachten will. Die literaturgeschichtlichen Ambitionen der Sammlung erfüllen sich am ehesten über solche Formreihen und Gruppierungsversuche, die anfangen, Literaturgeschichte als Geschichte der Form zu praktizieren.⁶

2 Vorarbeiten: „Stylometry beyond attribution“⁷

Die Orientierung der Herausgeber an Gruppierungen und Formreihen nach dem Kriterium ‚Stil‘ und ihr Bewusstsein für die Ähnlichkeit und Distinktivität von Texten innerhalb von Gruppen und großen Textmengen lässt die Idee des Distanzmaßes auch im Sinne einer messbaren Größe als angemessen und für die statistische Korpusanalyse geeignet erscheinen. Wie viele auf dem Gebiet der literaturwissenschaftlichen Stilometrie haben wir dafür das R-package ‚Stylo‘ verwendet, die Daten für die Modellierung aber in das Netzwerkanalysetool ‚Visone‘ exportiert.⁸ Verarbeitet haben wir dabei die Distanzmatrix, die man sich in Stylo ausgeben lassen kann. Sie enthält die Distanzen zwischen allen 86 Novellen,

(Hg.), *Archiv/Fiktionen. Verfahren des Archivierens in Literatur und Kultur des langen 19. Jahrhunderts*, Freiburg i. Br. 2016, 385–408.

⁵Paul Heyse/Hermann Kurz (Hg.), *Der Deutsche Novellenschatz*, 24 Bde., München/Leipzig 1871–1876, hier: Bd. 1, 20.

⁶Die literaturgeschichtliche Stellung des *Novellenschatzes* zwischen Positivismus und Formalismus habe ich an anderer Stelle herausgearbeitet: vgl. Thomas Weitin, „Literarische Heuristiken. Die Novelle des Realismus“, in: Albrecht Koschorke (Hg.), *Komplexität und Einfachheit. DFG-Symposium 2015*, Stuttgart 2017, 422–441.

⁷Maciej Eder, „Visualization in stylometry. Cluster analysis using networks“, in: *Digital Scholarship in the Humanities* 32/1 (2017), 50–64, hier: 53.

⁸Siehe www.visone.info (letzter Aufruf 4.5.2018).

berechnet für eine bestimmte Wortlistenlänge auf der Basis eines festzulegenden Distanzmaßes, wobei wir mit dem klassischen *Burrows's Delta* gearbeitet haben, also mit den z-transformierten relativen Frequenzen. Berücksichtigt worden sind die 500 häufigsten Wörter einschließlich der Pronomen.

Für das Netzwerk, das Abb. 1 zeigt, ist die Distanzmatrix um einen zusätzlichen Vektor ergänzt worden, der für jeden Text die Distanz zum Gesamtkorpus enthält. Dazu haben wir alle Texte in eine Datei gegeben und für jeden Einzeltext den Deltawert berechnet. Wir haben dann den Deltawert für das Verhältnis je zweier Texte zueinander auf die Kanten und den Deltawert für das Verhältnis jedes einzelnen Textes zum Gesamtkorpus auf die Knoten geschrieben. Kanten und Knoten wurden nach den Deltawerten so skaliert, dass helle Färbung einen großen, dunkle einen kleinen Wert markiert. Dunkle Knoten und Kanten signalisieren mithin größere, hellere geringe Ähnlichkeit, und zwar einerseits im Verhältnis der einzelnen Texte zueinander und andererseits im Verhältnis des Einzeltextes zur Gesamtheit aller Texte. Unsere Überlegung dabei war, dass dieses von uns so bezeichnete *Korpusdelta* auf den Knoten ein stilometrischer Wert für die Distinktivität des einzelnen Textes ist, dessen intertextuelle Relationen durch das paarweise Delta auf den Kanten abgebildet wird.¹⁰ Außerdem wollten wir untersuchen, welche Gruppen sich auf dieser Basis innerhalb des Netzwerks identifizieren lassen. Dazu haben wir einen nach dem Soziologen Georg Simmel *Simmelian Backbone* genannten Algorithmus zur Transformation des Netzwerks genutzt. Er wurde ursprünglich für die Analyse sozialer Netzwerke entwickelt und hat das Ziel, untergründige Gemeinschaftsstrukturen herauszufiltern. Dazu werden Kriterien definiert, welche Verbindungen im Gesamtnetzwerk beibehalten werden sollen und welche nicht. Die zugrunde gelegte abstrakte Vorstellung von der Emergenz eines Netzwerks stammt aus der soziologischen Theorie Simmels und geht davon aus, dass die Beziehung zwischen Zweien immer erst in Bezug auf einen Dritten verständlich wird, der umgekehrt von der dyadischen Beziehung abhängt.¹¹ Eine Verbindung von B zu C ist wahrscheinlicher, wenn zwischen A und B und A und C auch eine Verbindung besteht: „Without dependence among ties, there is no emergent network structure.“¹² Wer Netzwerke nutzen will, um Beziehungen zu verstehen, sollte sie demnach nicht einfach aus Dyaden aggregieren. Der *Simmelian Backbone* sucht stattdessen nach Verbindungen innerhalb von Dreiecken – nach Kanten also, die stark eingebettet sind und innerhalb

¹⁰Vgl. Thomas Weitin, „Average and Distinction. The *Deutsche Novellenschatz* between Literary History and Corpus Analysis“, in: *LitLab Pamphlet* 6 (März 2018), <http://www.digitalhumanitiescooperation.de/pamphlete/> (letzter Aufruf 4.5.2018). Der Artikel erläutert ausführlich eine Vorstufe dieses Modells.

¹¹Vgl. Ulrik Brandes/Bobo Nick/Conrad Lee u. a., „Simmelian Backbones: Amplifying Hidden Homophily in Facebook Networks“, in: *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (2013), 525–532, hier: 525 f.

¹²Ulrik Brandes/Garry Robins/Ann McCraine u. a., „What is Network Science?“, in: *Network Science* 1 (2013), 1–15, hier: 10.

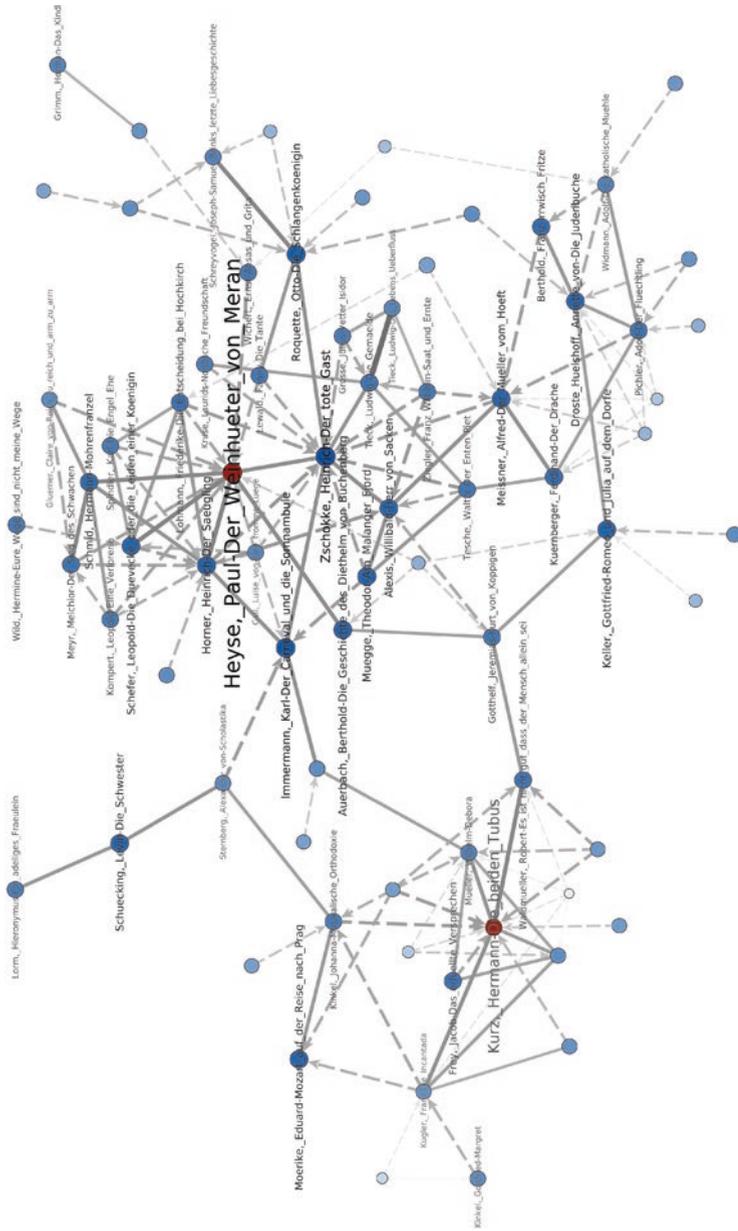


Abb. 1 Simmelian-Backbone-Netzwerk des Deutschen Novellenschatzes (Modell 1)⁹

⁹Die Farbstufen der Knoten, Kanten und Label sowie die Delta-Werten skaliert (500 MFW, Burrows-Delta, 20 % Culling). An den Kanten liegt der Delta-Wert für das paarweise Verhältnis je zweier Texte zueinander an, an den Knoten der Wert für das Verhältnis jedes einzelnen Textes zum Gesamtkorpus.

von Umgebungen vorkommen, in denen viel Redundanz herrscht. Kanten, die nur schwach eingebettet sind, werden herausgefiltert, es sei denn, ohne sie würde das Netzwerk auseinanderfallen. Daher die Bezeichnung *Backbone*. Operationalisiert wird das innerhalb von Visone, indem *Overlaps* in den Nachbarschaftsbeziehungen definiert werden. In unserem Modell haben wir festgelegt, dass für jeden Text die Rangfolge von sechs Nachbarn in der geringsten Delta-Distanz berücksichtigt wird und dass nur diejenigen Verbindungen zwischen Texten erhalten bleiben sollen, die drei gemeinsame Nachbarn haben.

Was wir damit erreichen, ist ein lokal adaptives Filtern, das das Netzwerk so zum *Backbone* transformiert, dass wir einzelne Gruppen und die *Weak Links* zwischen ihnen identifizieren können. Weil wir im *Overlap* den *Delta Rank* aufsteigend vom kleinsten Wert definiert haben und mithin die geringsten Distanzen zugrunde legen, sind es lokale Ähnlichkeitsgruppen, die sich herausgebildet haben. Der *Backbone* des *Novellenschatzes* (s. Abb. 1) zeigt uns links eine klar unterschiedene Untergruppe, in welcher der Herausgeber Hermann Kurz einen zentralen Knoten innehat. Eine andere, größere Gruppe ist um den zweiten Herausgeber Heyse sowie Zschokke und Horner entstanden. Ob die Verbindungen unten rechts um Droste und Meißner zu dieser Gruppe zu zählen oder als eigenes Teilnetzwerk zu verstehen sind, bleibt unklar.

In der Interpretation dieses Modells haben wir uns auf die Ähnlichkeitsgruppen sowie auf die Ähnlichkeitsverhältnisse konzentriert, die durch die Skalierung der Knoten und Kanten nach den Deltawerten sichtbar sind. Auf den Kanten liegt das paarweise Delta. Dunkle Kanten signalisieren einen kleinen Deltawert und mithin Ähnlichkeit zwischen den Texten, die sie verbinden. Auf den Knoten liegt das *Korpusdelta* an, der Abstand des Einzeltexts zur Gesamtheit aller Texte. Wer einen dunklen Knoten hat, der ist der Gesamtheit sehr ähnlich und mithin als ‚durchschnittlich‘ zu qualifizieren. Das gilt zum Beispiel für Heinrich Zschokke, Karl Immermann, Otto Roquette und Paul Heyse. Letzterer weist mit 0,516085434 das kleinste *Korpusdelta* im Korpus auf. Der zweite Herausgeber Hermann Kurz ist demgegenüber mit 0,700901735 deutlich weniger durchschnittlich (das Maximum ist 1,239941925), etabliert aber ein starkes lokales Ähnlichkeitszentrum, sichtbar durch die zahlreichen bei ihm eintreffenden gerichteten Kanten, die durch Strichelung hervorgehoben sind. Beim Überlappen (*Overlap*) der Nachbarschaftsbeziehungen, das dem *Backbone* zugrunde liegt, gibt es Asymmetrien. Es gibt Texte, die in den *Top Ranks* der Ähnlichkeit von sehr vielen Texten vorkommen, ohne dass das umgekehrt auch der Fall ist. Diese Texte sind für sehr viele andere Texte diejenigen, mit denen sie die meisten gemeinsamen Nachbarn haben. Die gestrichelten gerichteten Kanten weisen auf solche Texte hin, die einen anderen Typus von Ähnlichkeit darstellen als den der globalen Durchschnittlichkeit. Wir können demnach zwischen Texten unterscheiden, die global sind, und solchen, die lokal ähnlich sind, wodurch uns unsere beiden Herausgeber als unterschiedliche Ähnlichkeitstypen erscheinen. Hermann Kurz, der weniger durchschnittliche, ist innerhalb seiner Subgruppe ähnlich. Paul Heyse ist sehr durchschnittlich und zugleich für viele Texte naher Nachbar. Natürlich ist die Versuchung groß, dieser Beobachtung hermeneutischen Sinn zu verleihen, wenn man etwa an

die biografischen Unterschiede zwischen dem schwäbischen Dorfgeschichtendichter und dem seine Epoche prägenden Großschriftsteller denkt, der zu fast allen irgendwie bedeutenden Kolleginnen und Kollegen Verbindungen hatte und unbestritten eine „Schlüsselfigur im literarischen Netzwerk“¹⁵ war. Hat dieser ob seiner marktgerechten Massenproduktion von Novellen oft auch verspottete „Virtuose des Durchschnitts“¹⁶ für den *Novellenschatz* ausgewählt, was zu ihm passte, und ist deshalb auch stilometrisch ähnlich mit allen? Die bislang beste Kennerin des Herausgeber-Briefwechsels Heyse/Kurz, Monika Walkhoff,¹⁷ widerspricht mit ihrer Darstellung jedenfalls der Idee, Heyse habe sich vom eigenen Geschmack leiten lassen, und führt viele negative Urteile über aufgenommene Texte an, die sich auch in den kurzen Einleitungen innerhalb des *Novellenschatzes* niederschlagen. Wie immer, wenn wir statistische Ergebnisse hermeneutisch interpretieren, geraten wir mit solchen Überlegungen ohnehin leicht in den Bereich der Spekulation.

3 Was macht den Unterschied aus?

3.1 Distanzmatrizen

Dies umso mehr, als wir das Netzwerkmodell vor allem genutzt haben, um ein Gefühl für das Korpus und seine Daten zu bekommen. Dass in der Novellensammlung große Ähnlichkeit herrscht, konnte uns angesichts der von den Schriftstellerinnen und Schriftstellern stark empfundenen und diskutierten ‚Novellenflut‘ in der zweiten Hälfte des 19. Jahrhunderts nicht überraschen. Im *Novellenschatz* schlägt sich das offenbar auch auf der Ebene nieder, die man mit stilometrischen Methoden erfasst. Wir haben uns die Distanzmatrizen, die Stylo mit ausgibt, daraufhin eingehend angeschaut und uns zunächst auch vom Eindruck der Ähnlichkeit leiten lassen. Mit sehr geringem Aufwand lässt sich die Distanzmatrix, die die paarweisen Delta-Abstände für alle 86 Novellen enthält (s. Tab. 1), um die deskriptive Statistik ergänzen, die uns die zentrale Tendenz dieser Werte erkennen lässt.

Weil nur vier Autorinnen und Autoren mit je zwei Novellen in der Sammlung vertreten sind, benennen wir die Texte nur noch mit den Nachnamen und geben ausschließlich bei den Doppelaufträgen zusätzlich den Novellenkurztitel an.

¹⁵ Urzula Bonter, *Paul Heyse. Hofdichter und Publikumsschriftsteller*, in: Krzysztof Ruchniewicz/ Marek Zybura (Hg.), *Die höchste Ehrung, die einem Schriftsteller zuteil werden kann. Deutschsprachige Nobelpreisträger für Literatur*, Dresden 2007, 61–88, hier: 74.

¹⁶ Klaus Jeziorkowski, „Der Virtuose des Durchschnitts. Der Salonautor in der deutschen Literatur des 19. Jahrhunderts, dargestellt am Beispiel Paul Heyse“, in: Ders., *Eine Iphigenie rauchend. Aufsätze und Feuilletons zur deutschen Tradition*, Frankfurt a. M. 1987, 114–129, hier: 114.

¹⁷ Monika Walkhoff, *Der Briefwechsel zwischen Paul Heyse und Hermann Kurz in den Jahren 1869–1873 aus Anlass der Herausgabe des Deutschen Novellenschatzes*, München 1967.

Tab. 1 Ausschnitt Alexis-Eichendorff aus der Distanzmatrix des Novellenschatzes¹³

	Alexis	Andolt	Arnim	Auerbach	Berthold	Brentano	Chamisso	Dincklage	Droste	Eichen- dorff
Alexis	0,0000	0,8715	0,9812	0,8362	0,8643	1,0694	0,9537	0,9230	0,8737	1,0844
Andolt	0,8715	0,0000	1,0411	1,0087	1,0340	1,1589	0,9763	1,0453	0,9806	1,1546
Arnim	0,9812	1,0411	0,0000	0,9378	1,0458	1,0876	1,0685	1,1002	1,0578	1,1322
Auerbach	0,8362	1,0087	0,9378	0,0000	0,8386	1,0358	0,9859	1,0219	0,8306	1,0120
Berthold	0,8643	1,0340	1,0458	0,8386	0,0000	1,0106	0,9154	0,9784	0,7988	0,9518
Brentano	1,0694	1,1589	1,0876	1,0358	1,0106	0,0000	1,1394	1,1159	1,0934	1,2263
Chamisso	0,9537	0,9763	1,0685	0,9859	0,9154	1,1394	0,0000	1,1214	0,9872	1,1167
Dincklage	0,9230	1,0453	1,1002	1,0219	0,9784	1,1159	1,1214	0,0000	0,9506	1,1541
Droste	0,8737	0,9806	1,0578	0,8306	0,7988	1,0934	0,9872	0,9506	0,0000	0,9764
Eichendorff	1,0844	1,1546	1,1322	1,0120	0,9518	1,2263	1,1167	1,1541	0,9764	0,0000

¹³ Burrows-Delta, 500 MFW, 20 % Culling.

Tab. 2 Die zehn kleinsten Werte für Mean und Median der paarweisen Delta-Distanz, aufsteigend sortiert¹⁴

	Mean		Median
Heyse	0,9003302	Heyse	0,8896433
Immermann	0,9271726	Immermann	0,9086326
Roquette	0,9304626	Moerike	0,9167870
Moerike	0,9311495	Zschokke	0,9196800
Alexis	0,9331512	Schuecking	0,9214096
Schuecking	0,9372467	Roquette	0,9219233
Zschokke	0,9392342	Alexis	0,9224916
Meißner	0,9414294	Meißner	0,9316439
Muegge	0,9496409	Horner	0,9400061
Horner	0,9534774	Muegge	0,9404860

Der Data-Viewer von *R-Studio* verfügt über eine effektive Sortierfunktion, dank welcher die Matrix nach jedem Vektor sortiert werden kann. Wenn ich der Matrix die Werte für die deskriptive Statistik als Vektoren hinzufüge, kann ich auch nach diesen Vektoren sortieren. Tab. 2 zeigt jeweils die zehn Novellen mit dem niedrigsten Durchschnittswert und Median aller paarweisen Deltadistanzen. Das *Mean Delta* (durchschnittlicher Delta-Abstand) verwenden wir inzwischen anstelle des *Korpusdelta* als Richtwert für die Distinktivität/Korpusähnlichkeit des einzelnen Textes.

Heyse hat das kleinste *Mean Delta* und auch den kleinsten *Median*. Wir erinnern diese Top Ten der Durchschnittlichkeit als dunkle Knoten in unserem anfänglichen Netzwerkmodell (s. Abb. 1). Dass Heyse besonders ähnlich mit dem Korpusdurchschnitt ist, zeigte sich uns sehr anschaulich in zwei weiteren Experimenten. Wir haben das Gesamtkorpus als zusätzliche Datei mit den 86 Novellen in eine Clusteranalyse gegeben, die Heyse als nächsten Nachbarn der Gesamtdatei zuordnete (s. Abb. 2).

Wir haben dieses Experiment mit einer schrittweise von 10 bis zu 4000 MFW erhöhten Wortlistenlänge wiederholt, wobei Heyse durchgehend in der unmittelbaren Umgebung des Gesamtkorpus clusterte. Bis zu einer Wortlistenlänge von 1000 MFW war er fast immer der nächste Nachbar der *Novellenschatz*-Gesamtdatei, danach rückte Auerbach ebenso konstant in diese Position (gefolgt von Heyse), während Zschokke mit zunehmender Wortlistenlänge aus diesem Ähnlichkeitscluster herausrückte.

Das zweite Experiment arbeitet wiederum mit einem Netzwerkmodell, wobei wir dieses Mal nicht die *Backbone*-Transformation, sondern ein *Centrality Layout* verwendet haben (s. Abb. 3).

¹⁴Burrows-Delta, 500 MFW, 20 % Culling.

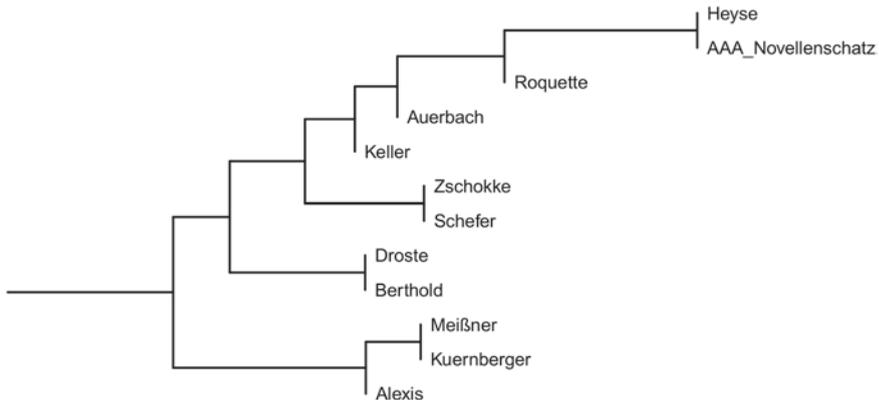


Abb. 2 Clustering aller Novellen mit dem Gesamtkorpus¹⁸

Auf den Knoten liegt das *Mean Delta*, gemappt auf die Labelgröße. Wir haben die Werte invertiert, sodass ein großes Label ein kleines *Mean Delta* und also hohe Korpusähnlichkeit anzeigt. Das *Centrality Layout* stellt den Knoten mit dem (durch die Invertierung) größten Wert in die Mitte. Die herausragende Korpusähnlichkeit Heyses kann so im Vergleich zu den anderen Texten abgelesen werden. Auf den Kanten liegt die (ebenfalls invertierte) paarweise Delta-Distanz, gemappt auf die Farbe. Wir haben hier zusätzlich einen Schwellenwert von $>0,775$ als Filter definiert, um nur die Verbindungen zwischen Texten zu zeigen, die sich besonders ähnlich sind. Dementsprechend zeigen sich bei Heyse viele und ausschließlich dunkle Kanten, während die Texte auf der ‚äußeren Umlaufbahn‘, deren invertiertes *Mean Delta* klein ist, wenige, deutlich hellere Kanten haben. Auf diese im Korpus besonders unähnlichen Texte – von Wolf, Wallner und Scheffel – werden wir noch zurückkommen.

3.2 Diskussion

In der Diskussion unserer Analysen kristallisierten sich schon bald zwei naheliegende Kritikpunkte heraus. Wir begannen zu hinterfragen, inwiefern Durchschnittswerte wie *Korpusdelta* oder *Mean Delta* überhaupt geeignet waren, unsere Novellen zu vergleichen. Zwar lässt sich die besondere Anfälligkeit des arithmetischen Mittels für Ausreißer durch den Abgleich mit dem diesbezüglich robusteren Median einholen. Trotzdem hat dieser Vergleichswert ein systematisches Problem, wenn man Ähnlichkeiten oder gar Ähnlichkeitstypen auf der Basis von Durchschnitten feststellen will. Wenn ich ein Ding mit dem Wert

¹⁸Burrows-Delta, 500 MFW, 20 % Culling (Ausschnitt).

Unser zweiter Kritikpunkt betraf die Konzentration auf die Ähnlichkeit selbst, die sich zweifellos als Leithypothese aus dem historischen Kontext unseres Korpus ableiten lässt und die Sammlung als Artefakt gut charakterisiert, zu der aber gerade in der literarischen Kommunikation der Entstehungszeit und explizit auch in der Gattungstheorie des *Novellenschatzes* die Suche nach Distinktionsstrategien gehört. Für die moderne Vorstellung von ‚Stil‘ sind die individuellen Unterschiede konstitutiv. Uns fiel in diesem Zusammenhang die alte Statistik-Geschichte von Hund und Krokodil ein, die die Erklärungskraft von Variablen veranschaulichen soll. Hund und Krokodil haben viele Ähnlichkeiten, die man gut beschreiben kann. Beide sind Vierbeiner, können laufen und schwimmen, haben vorn das Maul und hinten einen Schwanz usw. All das ist richtig, aber ein einziger effektiver Unterschied erklärt das Verhältnis viel besser: ‚Säugetier/Reptil‘.²¹ Auf den *Novellenschatz* bezogen heißt das: Die Ähnlichkeit im Korpus und die unterschiedliche Art und Weise, in der Texte ähnlich erscheinen, sind durchaus interessante Fakten. Aber es fällt uns sehr schwer zu sagen, was genau diese Ähnlichkeit ausmacht und wodurch sie entsteht. Zwar können wir uns dank der Matrizen mit den relativen Häufigkeiten bestimmte Wörter anschauen und Gemeinsamkeiten in der Häufigkeit zwischen verschiedenen Texten feststellen. Wenn aber viele Wörter und viele Texte im Spiel sind, können wir kaum sicher sein, dass tatsächlich (nur) diese Gemeinsamkeiten die stilometrische Ähnlichkeit begründen. Auch aufwendig, aber doch viel leichter ist es dagegen zu untersuchen, welche Wörter für Unterschiede verantwortlich sind und Texte unähnlich machen.

3.3 Frequenzmatrizen: Welche Wörter machen global einen starken Unterschied zwischen Texten?

In diese Richtung haben wir uns zunächst einen sehr simplen Schritt überlegt. Wenn man die Matrix mit den relativen Wortfrequenzen, die Stylo ausgibt, so transponiert, dass die Wörter die Spalten und die Textnamen die Reihen bilden, kann man wie bei der Distanzmatrix für jede Spalte, also für jeden Wortvektor, die deskriptive Statistik berechnen. Durch die kompakte *summary()*-Funktion macht es uns R sehr leicht, für die Häufigkeit eines jeden Worts in den 86 Texten alle Maße der zentralen Tendenz (*Mean*, *Median* etc.) zu betrachten. Uns kam es hier allerdings vor allem auf ein Maß der Dispersion an, nämlich auf die Standardabweichung der Wortvektoren.

Die Standardabweichung gibt an, wie stark die Werte in einem Datensatz ‚streuen‘ – oder besser gesagt: wie stark sie vom Mittelwert abweichen. Für jeden unserer Wortvektoren (s. Tab. 3) ist das ein Maß dafür, wie stark sich die Häufigkeitswerte des Wortes in den einzelnen Novellen vom Durchschnitt unterscheiden. Eine hohe Standardabweichung heißt, dieses Wort ist besonders diskriminativ

²¹ Ikira Schielke (Darmstadt) hat dieses Beispiel in die Diskussion eingebracht.

Tab. 3 Ausschnitt Alexis-Goethe (row) und der 10 häufigsten Wörter (column) der Frequenzmatrix des Novellenschatzes²⁰

	Und	Die	Der	Sie	Er	Zu	In	Ich	Den	Das
Alexis	2,6208	2,9664	3,1370	1,0807	2,3495	1,4569	1,2951	0,8663	1,3826	0,9407
Andolt	2,5122	2,0026	2,4766	1,5695	1,0192	1,6052	1,6358	3,0218	1,1567	1,3912
Arnim	3,4404	2,5517	2,3359	1,6758	2,3613	1,6250	1,5361	1,3457	1,5488	1,0537
Auerbach	4,0927	2,4071	2,5780	1,1548	2,6115	1,2615	1,2875	0,7703	1,0876	1,1563
Berthold	3,1749	2,7796	2,6038	2,9193	1,6294	1,4816	1,2141	0,8826	1,2620	1,0863
Brentano	3,8920	2,4406	2,7730	1,8325	2,4001	1,1676	1,2162	2,6433	1,2487	1,1514
Chamisso	3,5164	2,1838	1,9862	1,2465	1,2870	1,7278	1,3326	4,2714	1,4289	0,9576
Dincklage	3,3298	3,1845	1,9424	1,5856	0,5285	0,9249	1,2817	1,0042	1,1232	1,0835
Droste	3,2401	2,2959	2,6045	1,1479	1,7466	1,1171	1,2097	0,6851	1,4812	1,1788
Eichendorff	3,9378	3,1434	2,9983	1,4922	1,6649	0,7668	1,5199	0,4145	1,7962	1,0984
Ense	3,4222	2,6972	1,4746	2,0214	0,5530	1,8063	1,5237	2,8017	0,9339	1,0199
Fraulein Wolf	2,5252	1,6069	1,5566	0,9798	1,2990	1,6573	1,3830	2,4300	0,7055	0,7951
Frey	2,7663	2,5782	2,8122	0,7524	1,7341	1,4588	1,3717	0,8166	1,1790	1,2524
Gall	3,1012	2,6624	2,6185	2,3552	1,4555	1,5140	0,9508	0,8850	1,2068	1,3166
Gerstaecker	4,3994	2,4255	2,3205	1,0605	1,7850	1,0815	1,2810	0,9345	1,4070	1,7010
Gluemer	3,3708	2,7348	2,8408	1,9575	1,6465	1,4133	1,0247	1,6536	1,0671	1,2579
Goethe	3,9255	2,0760	1,2582	2,1137	0,4026	2,0634	1,3966	4,0513	0,8052	0,8556

²⁰Burrows-Delta, 500 MFW, 20 % Culling.

Tab. 4 Die 10 Wörter innerhalb der 86 Novellen mit der höchsten Standardabweichung, absteigend sortiert

Wörter	Standardabweichung
Ich	1,00173313
Sie	0,75733176
Und	0,66031381
Der	0,57509821
Er	0,52492546
Die	0,38317366
Mir	0,31514919
Mich	0,31352642
Es	0,29220622
Zu	0,27648317

und macht einen starken Unterschied zwischen den Texten. Tab. 4 zeigt die zehn Wörter unserer 86 Novellen mit der höchsten Standardabweichung.

Wie wir sehen, handelt es sich ausschließlich um Pronomen und Funktionswörter, wobei das Personalpronomen der ersten Person Singular die herausragende Spitze bildet und seinen Dativ und Akkusativ ebenfalls unter die Top Ten reiht. In der grafischen Darstellung aller Standardabweichungen für die 500 häufigsten Wörter können wir die Beobachtung der Spitzengruppe einordnen und mit der Gesamtheit vergleichen (s. Abb. 4).

Die Zipf'sche Verteilung zeigt, dass die Funktionswörter, die knapp bis zum Wortlistenrang 100 reichen, nicht nur eine höhere Standardabweichung aufweisen, sondern sich untereinander in ihrer Standardabweichung auch stärker unterscheiden als die auf der flachen Restkurve liegenden Inhaltswörter. Funktionswörter sind besonders diskriminativ, und sie variieren der Streuung im steilen Kurventeil nach zu urteilen innerhalb ihrer Kohorte besonders stark. Bestimmte Funktionswörter, in unserem Korpus z. B. das Personalpronomen ‚ich‘, etablieren

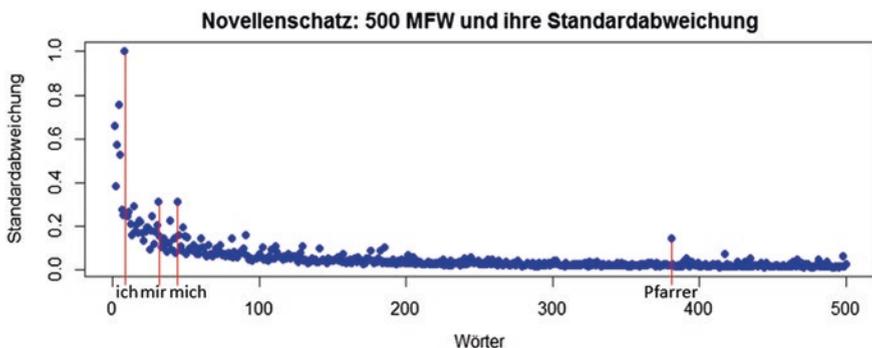


Abb. 4 Die 500 häufigsten Wörter des Novellenschatzes nach ihrer Standardabweichung

also einen herausragenden Unterschied zwischen Texten. Wir sehen unter den Inhaltswörtern ebenfalls Ausreißer, die sich indes sehr einfach erklären. Das Wort ‚Pfarrer‘ hat eine deutlich erhöhte Standardabweichung, weil es in 54 der 86 Texte überhaupt nicht vorkommt.²²

Die Stilometrie muss naturgemäß an den hochdiskriminativen Wörtern ein spezielles Interesse haben, und seit Burrows's Arbeiten zur Autorschaftsattribuion stehen dementsprechend die Funktionswörter im Fokus. Allerdings hat Maciej Eder zuletzt darauf aufmerksam gemacht, wie wenig Klarheit wir nach wie vor darüber haben, welche Merkmale dadurch tatsächlich unterschieden werden. Die theoretisch begründete Annahme, Funktionswörter unterschieden Autorschaft, Inhaltswörter Gattung und Topics, ist in der empirischen Praxis durch verschiedene Ergebnisse herausgefordert worden.²³ Eder verweist darauf, dass gerade die sehr diskriminativen Wörter mal für dieses, mal für jenes Signal reklamiert worden seien, der Artikel ‚the‘ etwa sowohl als Gendermarker²⁴ als auch als Unterschied zwischen britischem und amerikanischem Englisch.²⁵ Mit Blick auf die 86 Texte des *Novellenschatz*-Korpus' können wir zumindest klar sagen, was das diskriminativste Wort unterscheidet. Abb. 5 zeigt die relativen Frequenzen für das Personalpronomen der ersten Person Singular im Nominativ, Dativ und Akkusativ.

Wir haben probehalber einen Schwellenwert von 2 angenommen und uns angeschaut, welche Texte darüber liegen. Da die Novelle als ‚Schwester des Dramas‘ häufig unabhängig von der Erzählperspektive hohe Dialoganteile aufweist, waren wir sehr skeptisch, ob auf so simple Weise überhaupt ein erzählperspektivisch relevanter Unterschied zu ermitteln sein würde. Zumindest mit diesem hohen Schwellenwert gelang das aber schon. Unter den 18 ermittelten Texten befand sich nur einer mit ausgesprochen hohem Dialoganteil und ansonsten auktorialer Perspektive (Leo Goldammers *Eine Hochzeitsnacht* [1858]), alle anderen waren Ich-Erzählungen.²⁶ Als ein im Sinne des Spitzenwertes der Standardabweichung besonders diskriminatives Wort im Korpus unterscheidet das Pronomen ‚ich‘ also durchaus diese Perspektive.

²² Andere solche Wörter sind durch das Culling bereits aussortiert worden.

²³ Maciej Eder, „Visualization in stylometry. Cluster analysis using networks, in: *Digital Scholarship in the Humanities* 32/1 (2017), 50–64, hier: 53.

²⁴ James W. Pennebaker, *The Secret Life of Pronouns. What our Words Say About Us*, New York 2011, 42.

²⁵ Matthew Jockers, *Macroanalysis. Digital Methods and Literary History*, Chicago 2013, 105.

²⁶ August Wolfs *Der Stern der Schönheit* (1871) ist dabei eine ‚unechte‘ Ich-Novelle: Die Ich-Erzählung findet hier nur innerhalb der Binnenerzählung statt, wobei sowohl Rahmen- als auch Binnenerzählung dialogisch sind. Fräulein von Wolfs (Herausgeber-Kürzel ‚F.v.W‘) *Gemüth und Selbstsucht* (1837) ist eine Briefnovelle, die daher Ich-Erzählungen aneinanderreicht.

3.4 Matrixsubtraktion: Welche Wörter unterscheiden Texte im Einzelnen?

Nach diesem kleinen Test mit der Standardabweichung unserer Wortvektoren war uns klar, dass damit noch keine neue Basis für unser Textvergleichsmodell gefunden war. Wir mussten uns die Unterschiede durch einzelne Wörter nicht nur global, sondern im Einzelnen anschauen und untersuchen, welche Wörter im Einzeltextvergleich große Unterschiede ausmachen. Diese Wörter sollten die Datengrundlage für ein neues Netzwerkmodell sein, das von der Systematik her aufgebaut sein sollte wie das alte, also mit den paarweisen Unterschieden auf den Kanten und den Unterschieden zwischen jedem einzelnen Text und allen anderen Texten auf den Knoten, wobei der letzte Unterschied nun nicht mehr über das *Korpusdelta* (Text X-Gesamtdatei) oder das *Mean Delta*, sondern für jeden Text einzeln durch Matrixsubtraktion mit allen Texten zu errechnen war. Das trug der Kritik Rechnung, dass Durchschnittswerte nicht unbedingt gute Vergleichswerte seien. Einen Durchschnittswert stellt indes auch *Burrows's Deltawert* dar. Wenn er zwei Texte vergleicht, wird für eine definierte Wortlistenlänge für jedes Wort der beiden Texte der Z-Score-Differenz-Betrag gebildet und dann durch n , also durch die Länge der Liste, geteilt.²⁹ Als Durchschnittswert aller Z-Score-Differenz-Beträge stellt *Burrows's Delta* dann *einen* Wert dar, der angeben soll, wie ähnlich oder unähnlich zwei Texte sind. Uns ging es nun darum, unserem neuen Textvergleichsmodell eine höhere ‚Auflösungsstufe‘ zugrunde zu legen und anstelle der durchschnittlichen Distanz den größten Unterschied bzw. das Wort zu betrachten, das diesen Unterschied etabliert. Dazu sind wir im Vergleich zu Burrows's Vorgehen einen Schritt zurückgegangen und haben mit den Z-Scores gerechnet. Für die Kanten haben wir jeweils paarweise die Z-Score-Vektoren zweier Texte auf den größten Differenzbetrag hin verglichen, für die Knoten musste entsprechend für jeden einzelnen Text der Z-Score-Vektor mit der Matrix aller einzelnen 86 Vektoren auf das Maximum hin verglichen werden. Während wir also im ersten Textvergleichsmodell mit Burrows jeweils den Durchschnitt aller Z-Score-Differenz-Beträge der betrachteten Wörter als Distanzmaß zugrunde gelegt haben, stellt das zweite Modell den Vergleich über den größten Unterschied, das Maximum der Z-Score-Differenz-Beträge, dar.

Dabei haben wir wiederum unser ‚Simmel'sches Gerüst‘, also den *Simmelian Backbone*, als Gruppierungsalgorithmus angewendet. Simmels Soziologie entwickelt sich gegen Ende des 19. Jahrhunderts maßgeblich unter dem Eindruck eines wachsenden Distinktionsdrucks in der gleichförmigen Massengesellschaft sowie vor dem Hintergrund der Diskussion um das in der Massengesellschaft gleichfalls wachsende Bedürfnis nach bestimmten Gemeinschaften. Der *Simmelian Backbone* wurde angesichts dessen dafür konzipiert, Datenmengen so zu reduzieren, dass Gemeinschaftsstrukturen sichtbar werden. Ein Vergleich rein

²⁹Vgl. John F. Burrows, „Delta“. A Measure of Stylistic Difference and a Guide to Likely Authorship“, in: *LLC* 17 (2002), 267–287, hier: 270.

über maximale Unterschiede entspricht dieser Ausrichtung nur bedingt, weshalb wir uns überlegt haben, wie wir die Datengrundlage für unser zweites Modell so filtern, dass über die zugrunde gelegten Unterschiede auch Ähnlichkeitsgruppen entstehen können. Im Ergebnis haben wir den Algorithmus so eingestellt, dass wir nach Texten suchen, deren größter Unterschied (relativ) klein ist.³² Technisch kann man sich das so vorstellen, dass wir wie oben in der Distanz-Matrix von Tab. 1 für jeden Text einen Vektor mit 86 Distanzen zu den jeweils anderen Texten haben (mit dem Wert 0 im Selbstverhältnis auf der Diagonale), nur dass die Distanz nun nicht mehr durch den Delta-Wert ausgedrückt wird, der den Durchschnitt aller Z-Score-Differenzen erfasst, sondern durch denjenigen Z-Score-Wert, durch den sich die Texte jeweils maximal unterscheiden. Wie den Delta-Vektor können wir auch diesen Z-Score-Max-Vektor sortieren, in unserem Fall so, dass die kleinsten Maximalwerte oben stehen. Genau wie bei der Anwendung des Gruppierungsalgorithmus in Modell 1 stellen wir diesen so ein, dass wir für zwei Texte, die verglichen werden, jeweils sechs Nachbarn berücksichtigen. Das sind diejenigen sechs Texte, zu denen der maximale Unterschied, den ein bestimmtes Wort erzeugt, am kleinsten ist. Unsere zweite Bedingung lautet: Von diesen sechs Nachbarn müssen die beiden verglichenen Texte drei gemeinsam haben (*Overlap*), sonst werden die betreffende Verbindung und ihre beiden Knoten aus dem Netzwerk gelöst – sofern dieses dadurch nicht auseinanderfällt.

Bei diesem Modell (s. Abb. 6) fällt weniger eine Gruppierung als eine Zentrum-/Peripheriestruktur auf. Für die im Zentrum verbundenen Texte gilt, dass sie auch in ihrem größten Unterschied nicht stark differieren. Die Texte in der Peripherie unterscheiden sich dagegen im Extrem deutlicher, und die Texte unten, die vom Netzwerk gelöst worden sind, erfüllen die algorithmische Bedingung nicht mehr. Ihre größten Unterschiede können nicht mehr als klein gelten. Wir sehen sofort, dass der Heyse-Text erneut im Zentrum steht und um ihn herum einige derjenigen Texte, die zu seiner großen Ähnlichkeitsgruppe in Modell 1 gehört haben, etwa Fanny Lewald und Ludwig Tieck mit *Die Gemälde*. Andere dort zentrale Knoten wie Immermann, Alexis oder Roquette erscheinen dagegen in Modell 2 nur noch in der Peripherie oder sind wie Auerbach und Zschokke ganz aus dem *Backbone*-Netzwerk herausgefallen. Diese Unterschiede sind durch die je andere abstrakte Vorstellung von Ähnlichkeit bedingt, die wir den beiden Modellen zugrunde gelegt haben. Immermann und Zschokke etwa sind im Schnitt sehr ähnlich, was in Modell 1 durch den dunklen Knoten (kleines *Korpusdelta*) sichtbar ist. Beide gehören dementsprechend gemeinsam mit Heyse zur Top Ten der kleinsten *Mean Delta*-Werte (s. Tab. 2 u. 3). Nimmt man dagegen nicht die durchschnittliche Ähnlichkeit, sondern die Spitzen der Unähnlichkeit zum Kriterium, erscheinen beide unähnlich und gehören nicht mehr wie Heyse zum Ähnlichkeitszentrum von Modell 2, weil ihre großen Unterschiede wirklich (relativ) groß sind – bei Zschokke so groß, dass sie die Backbone-Anforderungen

³² Maßgeblichen Anteil an der Entwicklung dieses Netzwerkmodells hat Anastasia Pupynina (Darmstadt).

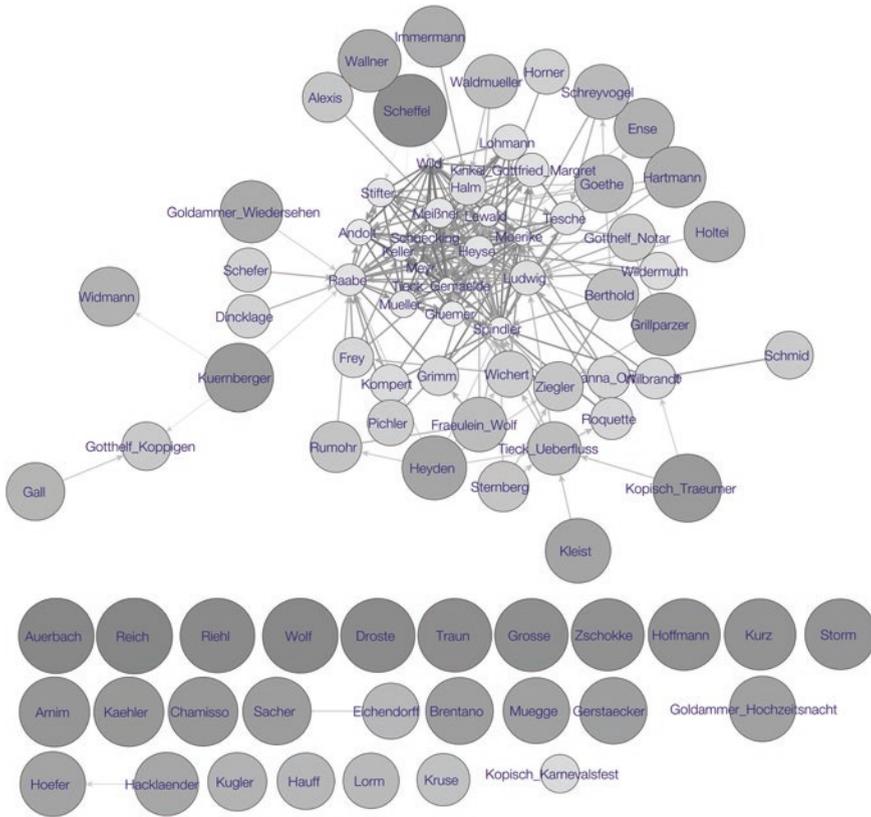


Abb. 6 Simmelian-Backbone-Netzwerk des Deutschen Novellenschatzes (Modell 2)²⁷

an Gemeinschaftlichkeit nicht mehr erfüllen. Dass das auch für den zweiten Herausgeber Kurz gilt, der ebenfalls unten als losgelöster Knoten erscheint, können wir im Modellvergleich recht gut verstehen. Wir hatten mit Blick auf Modell 1 gesagt, Kurz sei anders als Heyse nicht besonders durchschnittlich, der *Backbone* auf der Basis der Delta-Werte zeigt uns aber, dass er ein starkes lokales Ähnlichkeitszentrum ausbildet, dessen Unterschied zur Heyse'schen Hauptgruppe wir freilich mit dem *Korpusdelta* als Knotenwert nicht gut erklären können (s. Abb. 1). Die Kurz'sche Gruppe können wir zwar mit Modell 2 auch nicht

²⁷Die Farbstufe der Knoten und Kanten wurde nach dem maximalen Z-Score-Differenz-Betrag skaliert (Burrows-Delta, 500 MFW, 20 % Culling). An den Kanten liegt der maximale Z-Score-Differenz-Betrag für das paarweise Verhältnis je zweier Texte zueinander an, an den Knoten das Maximum, das für jeden Text im iterierten Einzelvergleich mit allen anderen Texten ermittelt wurde.

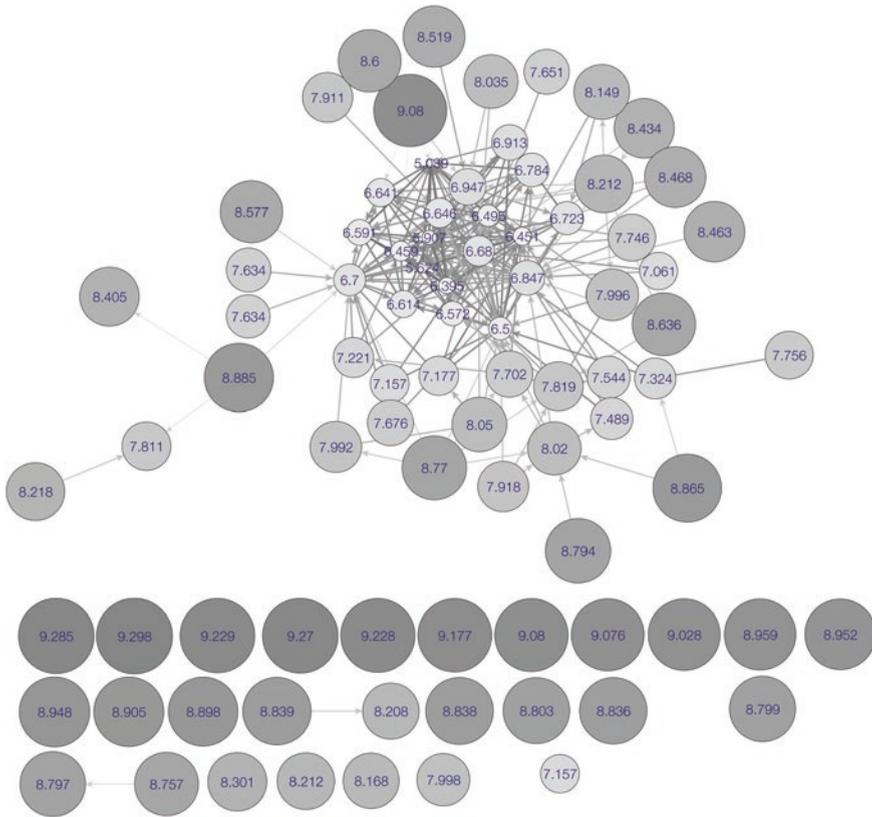


Abb. 7 Gleiches Netzwerk wie Abb. 6. Die Label zeigen nun die zugrundeliegenden Werte²⁸

deuten, aber dass Kurz aus dem neuen *Backbone*, in dem die großen Unterschiede klein sein müssen, herausfällt, führt uns doch zu einer weiteren Erkenntnis: Wir wissen jetzt, dass er nicht besonders durchschnittlich ist (Modell 1) und dass er sich im Extrem tatsächlich stark unterscheidet (Modell 2).

Wenn wir auf die Label der Knoten die Werte schreiben, aus denen das Netzwerk errechnet worden ist, wird noch etwas deutlich, das uns weiterhilft: Die Gruppierung in Modell 1 war beachtlich, hatte aber den entscheidenden Schönheitsfehler, dass wir die Gruppenunterschiede nicht gut beschreiben konnten. Was wir hier als Zentrum und Peripherie sehen, verhält sich anders. Der visuelle Eindruck des Modells wird durch die Werte bestätigt: Im Zentrum finden wir Texte, deren größter Unterschied durchweg klein ist, kein Wert geht hier über 7. In der deutlich

²⁸Für jeden Text wird der Z-Score-Differenz-Betrag desjenigen Wortes gezeigt, das ihn am stärksten unterscheidet. An den Kanten liegt entsprechend der Wert für den größten Wortunterschied zwischen zwei Texten.

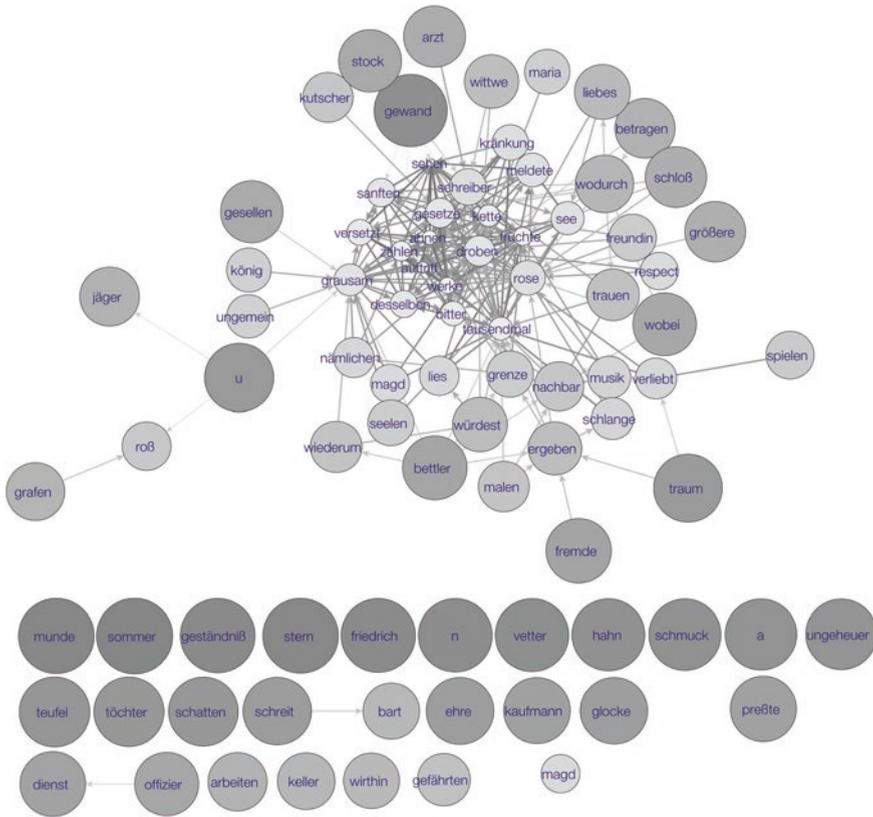


Abb. 8 Gleiches Netzwerk wie Abb. 6 und 7. Die Label zeigen nun das Wort, das jeden Text am stärksten unterscheidet³⁰

größeren Peripherie liegen dagegen alle Werte über 7 und reichen bis in den Bereich >9. Die aus dem Netzwerk herausgefilterten Knoten liegen mit nur einer deutlichen Ausnahme (Kopischs *Ein Karnevalsfest auf Ischia* [1856]) im Bereich >8 und >9. In diesem Modell sind wir also sehr wohl in der Lage, die Gruppenzugehörigkeit auf die von uns als Attribut definierte Eigenschaft zurückzuführen.

Diese Eigenschaft ist der größte Wortunterschied, wobei wir uns natürlich auch das Wort anzeigen lassen können, das diesen Unterschied jeweils ausmacht (s. Abb. 8).

Wir sehen dadurch, dass sich Kleists *Die Verlobung von St. Domingo* (1811) im Korpus am stärksten durch das Wort „fremde“ unterscheidet, Kopischs *Der*

³⁰An den Kanten liegt nach wie vor der Wert für den größten Wortunterschied zwischen zwei Texten.

Träumer (1856) recht naheliegend durch „traum“ und Raabes *Das letzte Recht* (1865) durch „grausam“. Die vielen bei diesem Wort eintreffenden Kanten kommen wie schon in Modell 1 durch eine Asymmetrie im *Backbone Overlap* zustande: Raabe gehört mit seinem distinktiven „grausam“ bei sehr vielen Texten zu den Top 6 der Wörter, für die der größte Unterschied zwischen den Texten klein ist. Er kommt also (wie Heyse mit seinem kleinen Delta-Wert als Zeichen sehr hoher Durchschnittlichkeit in Modell 1) in den *Top Ranks* von weit mehr als sechs Texten vor, sodass ein asymmetrisches Verhältnis entstehen muss, weil der Algorithmus unserer Einstellung nach für Raabe wie für jeden anderen nur die sechs nächsten Nachbarn berücksichtigt.³⁴

Die Kanten des Netzwerks sind nach dem größten Z-Score-Differenz-Betrag der beiden verbundenen Texte skaliert, also nach dem Wort, das paarweise den größten Unterschied etabliert. Auch diese Wörter können wir uns als Kantenlabel anzeigen lassen, was sich allerdings für das dichte Zentrum nicht übersichtlich ausskalieren lässt (Abb. 9).

Immerhin aber sehen wir in der Peripherie, dass „fremde“ auch das Wort ist, das Kleist bilateral im Verhältnis zu Tieck am stärksten unterscheidet, und dass es unter den sehr große Unterschiede etablierenden Wörtern jenseits des Netzwerkes der „mondschein“ ist, der Eichendorff und Sacher-Masoch trennt.

3.5 Diskussion

Wir sind von der Frage ausgegangen, welche Wörter dafür verantwortlich sind, dass Texte unähnlich erscheinen, und haben nach den größten Wort-Unterschieden gefragt, wobei das Gerüst unseres Netzwerks auf der zusätzlichen Bedingung beruht, dass wir nur Verbindungen von Texten berücksichtigen, die sich in den Extremen nicht stark unterscheiden. Diese Texte bilden das Zentrum des Netzwerks von Modell 2 (s. Abb. 6). Die Texte in der Peripherie genügen dieser Bedingung weniger, die abgehängten Texte gar nicht mehr. Wir haben gesehen, dass einige der in Modell 1 nach dem Delta-Verfahren als einander sehr ähnlich erscheinenden Texte aus der großen Gruppe um Heyse nun diesem Netzwerk nicht mehr angehören. Wenn wir nicht den Durchschnitt der Z-Score-Differenzbeträge für die gesamte Wortliste berechnen, sondern das Maximum zugrunde legen, ergeben sich für einige Texte – aber nicht für alle! – andere Verhältnisse.

³⁴Die Einzelbuchstaben auf den Knoten hielten wir zunächst für Korpusfehler, was aber nur in einem Fall („n“) zutrifft. Das sehr distinktive „a“, das Hermann Kurz' *Die beiden Tubus* (1859) am stärksten unterscheidet, stammt in diesem Fall von der für einen der beiden Titelhelden häufig gebrauchten geografischen Namensbezeichnung „Pfarrer von A“. In anderen Texten ist „a“ Ausdruck von dialektaler Rede – so in Heyses *Der Wein Hüter von Meran* (1864): „Daß ich dies Jahr möcht' werden a Braut“ – oder von Fremdsprache, z. B. in Mörikes *Mozart auf der Reise nach Prag* (1857): „À la bonne heure“.

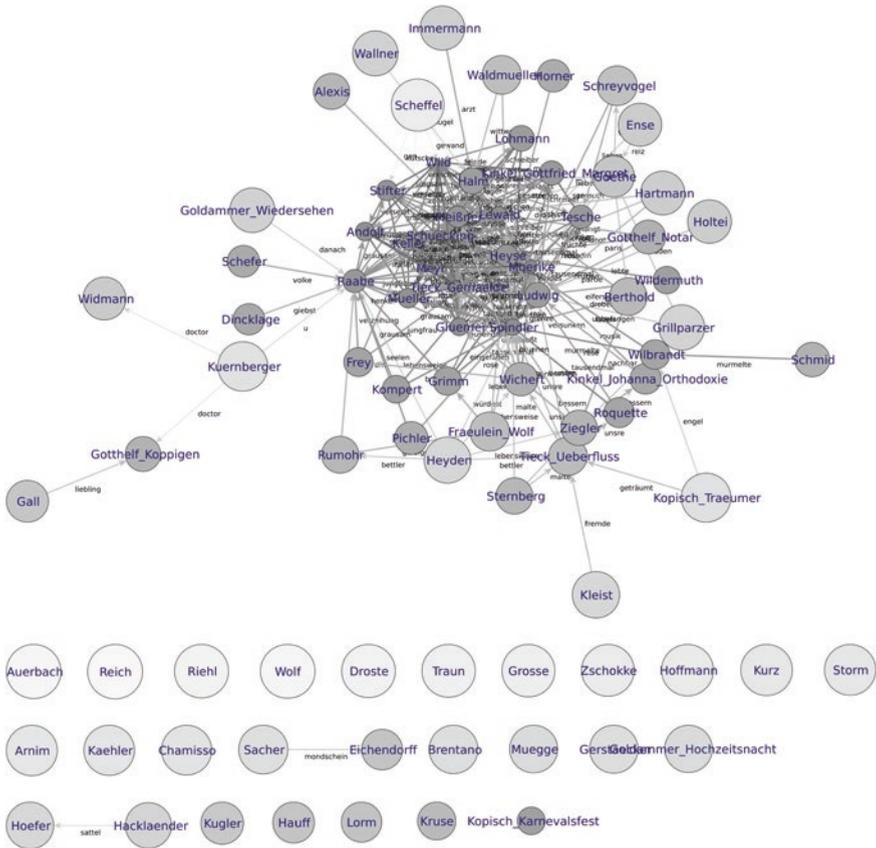


Abb. 9 Gleiches Netzwerk wie Abb. 6, 7 und 8. Dargestellt werden die Wörter, die beide Texte am stärksten unterscheiden³¹

Schauen wir uns die Wörter an, deren normalisierte Häufigkeit diese Verhältnisse etabliert (s. Tab. 4). In einem eigenen Experiment haben wir mit einer schrittweise von 10 auf bis 4000 MFW erweiterten Wortliste über den durchschnittlichen Delta-Abstand (*Mean Delta*) berechnet, welche Texte im Korpus am unähnlichsten sind. Ebenso regelmäßig wie Heyse, der „Virtuose des Durchschnitts“, das kleinste *Mean Delta* hat, weisen vier Texte die höchsten *Mean Delta*-Werte auf, sind also besonders unähnlich. Das sind August Wolfs *Der Stern der Schönheit* (1871), Victor von Scheffels *Hugideo* (1857), Franz Wallners *Der*

³¹Die Kanten-Label zeigen nun für den dort anliegenden Wert des größten Wortunterschieds, welches Wort es ist, das diese beiden Texte am stärksten unterscheidet.

arme Josy (1864) und Wilhelm Heinrich Riehls *Jörg Muckenhuber* (1861/62). Drei dieser vier Texte sind uns im *Centrality Layout* zur Korpusähnlichkeit (s. Abb. 3) bereits als Außenseiter aufgefallen. Wir haben dann über die oben beschriebene Matrixsubtraktion auf der Basis der Z-Scores ermittelt, welche Wörter es sind, die diese Texte so unähnlich sein lassen. Wir finden diese Wörter logischerweise fast ausschließlich unter denjenigen, die aus unserem „großer Unterschied klein“-Netzwerk in Modell 2 herausgefallen sind. Dazu gehören „munde“³⁸, „sommer“, „geständniß“, „stern“, „friedrich“, „vetter“, „schmuck“ und „hahn“. „Gewand“, das Wort, durch das sich Scheffels *Hugideo* im Korpus am stärksten unterscheidet, ist das erste dieser Wörter, das es in das Netzwerk von Modell 2 ‚schafft‘.

Den stark diskriminativen Wörtern stehen im Zentrum solche gegenüber, bei denen die Unterschiede zwischen den Texten weniger extrem sind. Wir sehen, dass wir es sowohl im Zentrum als auch am Rand und jenseits des Netzwerks mit Inhaltswörtern zu tun haben und dass auch jeweils das gleiche Spektrum von Wortarten in vergleichbarer Verteilung vertreten ist: Substantive, Adjektive, Adverbien und Verben. Der systematische Unterschied wird offensichtlich, wenn wir uns die Z-Score-Profile anschauen. „Sehen“ gehört zu den Wörtern im Zentrum mit den nicht so ausgeprägten Extremen. Entsprechend verteilen sich die Mittelwertabweichungen über verschiedene Zentren (s. Abb. 10 u. 11), wohingegen „sommer“, das nicht mehr zum Ähnlichkeitsnetzwerk von Modell 2 gehört, in Moritz Reichs *Mammon im Gebirge* (1858) einen einsamen Höhepunkt verzeichnet (s. Abb. 12). Wir sehen in diesem Z-Score-Profil an den zahlreichen identischen Negativabweichungen, dass „sommer“ in vielen Texten überhaupt nicht vorkommt, was auch bei der Mehrheit der Wörter mit weniger ausgeprägten Extremen der Fall ist. Stellvertretend dafür sei das Z-Score-Profil von „grausam“ mit angeführt (s. Abb. 10). Die Wörter im Zentrum, an der Peripherie und jenseits des Netzwerks unterscheiden sich also nicht dadurch, sondern durch die Verteilung der Mittelwertabweichungen über die Texte.

Fassen wir die Analysen zu unserem zweiten Textvergleichsmodell zusammen, können wir festhalten, dass es sowohl Funktions- als auch Inhaltswörter gibt, die viel diskriminativer sind als ihre ‚Artgenossen‘. Innerhalb der stilometrischen Methodendiskussion gehört diese Beobachtung zu der derzeit intensiv empirisch erforschten Frage, auf Basis welcher Wörter Texte am besten verglichen werden können. Eder berichtet beispielhaft von einer Clusteranalyse englischer Romane, bei der ein einziges unter 970 berücksichtigten Wörtern darüber entschied, ob ein Text von Joseph Conrad fälschlich Rudyard Kipling zugeschrieben wurde oder nicht: „The knowledge that this 970th word is ‚wine‘ does not help much, however, since multivariate analyses take into consideration a great number of

³⁸ „Munde“ ist deshalb so diskriminativ, weil es in Auerbachs *Die Geschichte des Diethelm von Buchenberg* (1853) als Namenskurzform (Raimund) einer der Hauptfiguren und mithin sehr häufig vorkommt.

Z-Scores von 'grausam' im Deutschen Novellenschatz

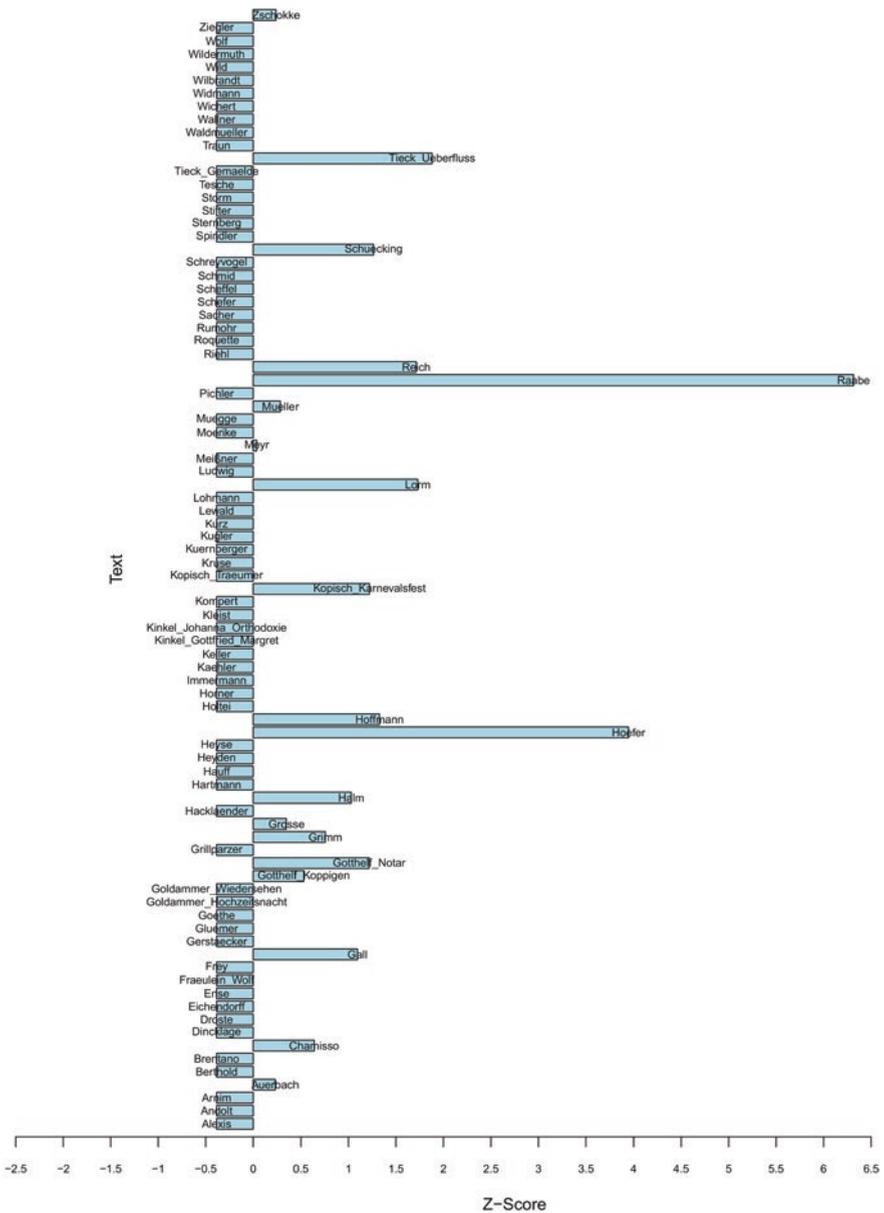


Abb. 10 Z-Score-Profil für „grausam“³³

³³Die Z-Score-Profile wurden aus dem nach der algorithmischen Bedingung „großer Unterschied klein“ gebildeten Zentrum von Netzwerkmodell 2 (siehe Abb. 8) hergestellt.

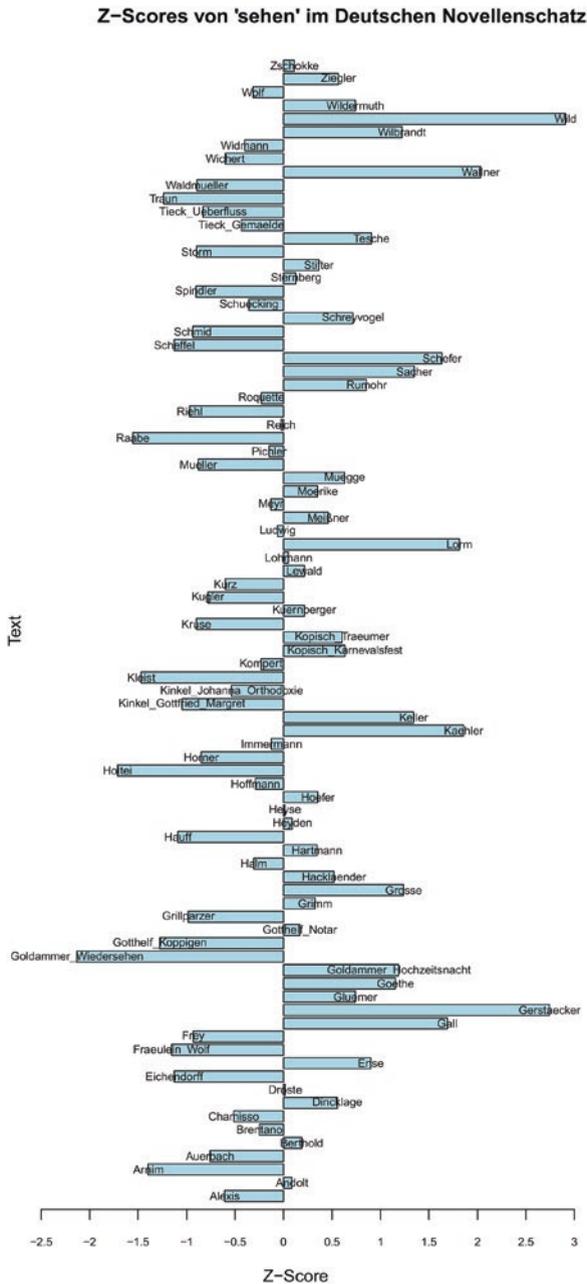


Abb. 11 Z-Score-Profil für „sehen“³⁵

³⁵ Die Z-Score-Profile wurden aus dem nach der algorithmischen Bedingung „großer Unterschied klein“ gebildeten Zentrum von Netzwerkmodell 2 (siehe Abb. 8) hergestellt.

der einzelnen Features und zieht deren Unterscheidungskraft in Betracht. Evert/Jannidis u. a.⁴⁰ schlagen nach der empirischen Auswertung des Erfolgs verschiedener Distanzmaße in der Autorschaftsattribuion⁴¹ ein Verfahren vor, das die Z-Score-Vektoren einem überwachten Preprocessing unterzieht, in dem u. a. diejenigen diskriminativen Wörter, die einen großen Unterschied ausmachen, weil sie in vielen Texten des untersuchten Korpus gar nicht vorkommen, aussortiert werden. Übrig bleibt ein deutlich kleinerer, normalisierter Vektor von diskriminativen Features, über den dann die Distanz bestimmt wird. In der Evaluation der verbliebenen diskriminativen Wörter, die sich im Sinne korrekter Klassifikation bewährt haben, betonen Evert/Jannidis, dass es sich beileibe nicht nur um Funktionswörter handle.⁴² Zudem vermuten die Autoren, dass sich Autorenstil primär in den positiven und negativen Abweichungen vom Korpusmittel niederschläge, wie sie die Z-Score-Vektoren repräsentieren.⁴³ Als wir untersucht haben, welche Wörter dafür verantwortlich sind, dass vier Texte des *Novellenschatzes* besonders unähnlich erscheinen, haben wir für jeden der Texte jeweils diesen Vektor zugrunde gelegt und Schritt für Schritt von allen Z-Score-Vektoren des Korpus subtrahiert. Durch Berechnung der Maxima und Minima der Ergebnisse dieser Matrixsubtraktion können wir je eine Rangfolge für die Wörter erstellen, die in den fraglichen Texten besonders überrepräsentiert (die höchsten positiven Werte) und besonders unterrepräsentiert (die niedrigsten negativen Werte) sind. In beiden Rängen finden wir Wörter, die schlechthin einen großen Unterschied ausmachen, weil ein Text ihr Vorkommen im Korpus fast im Alleingang bestreitet (s. Abb. 12: „sommer“). Angesichts solcher Dominanz stellt sich die Frage, inwiefern es denn charakteristisch für einen Text sein könne, dass in einem anderen Text ein Wort besonders häufig vorkommt. Evert/Jannidis ist der Nachweis gelungen, dass Wörter, die nur in relativ wenigen Texten eines Korpus vorkommen, im Schnitt deutlich weniger zur paarweisen Delta-Distanz beitragen.⁴⁴ Solche Wörter finden wir in unserem Modell 2 nicht nur in der Peripherie, wo sich die Texte im Extrem stark unterscheiden, sondern auch im Zentrum, wo die großen Unterschiede klein sind (s. Abb. 10: „grausam“). Dort aber handelt es sich um Wörter, die, obwohl sie in relativ vielen Texten nicht vorkommen, keine extremen Unterschiede etablieren. Die Mittelwertabweichungen

⁴⁰ Stefan Evert/Fotis Jannidis/Steffen Pielström u. a., „Towards a better understanding of Burrows’s Delta in literary authorship attribution“, in: *Proceedings of NAACL-HLT Fourth Workshop on Computational Linguistics for Literature* (2015), 79–88.

⁴¹ Fotis Jannidis/Steffen Pielström/Christof Schöch u. a., „Improving Burrows’ Delta – An empirical evaluation of distance measures“, in: *Digital Humanities Conference Sydney* (2015), https://www.researchgate.net/publication/280086768_Improving_Burrows%27_Delta_-_An_empirical_evaluation_of_text_distance_measures (letzter Aufruf 4.5.2018).

⁴² Evert/Jannidis/Pielström u. a. (Anm. 23), 86.

⁴³ Ebd., 84.

⁴⁴ Ebd., 83.

dieser Wörter verteilen sich auf eine relevante Textgruppe, sodass womöglich maßgebliche Beziehungen verloren gingen, würden sie für die Distanzberechnung aussortiert.

4 Ausblick: Entropie

Nachdem unsere beiden Textvergleichsmodelle die Distanz sowohl über den Durchschnittswert als auch über das Maximum der Z-Score-Differenzbeträge berechnet haben – das eine Mal auf Basis der gesamten definierten Wortliste, das andere Mal über bestimmte Einzelwörter –, schließt sich für weitere Untersuchungen logisch die Frage an, ob viele oder wenige Wörter für den Unterschied zwischen Texten verantwortlich sind. Für das *Novellenschatz*-Korpus sind wir dieser Frage mit einer Entropie-Berechnung nachgegangen.⁴⁵ Dabei wurde zunächst für alle Paarkonstellationen unserer 86 Texte auf der Basis des in beiden Modellen verwendeten Vektors von 500 MFW für jedes Wort die Wahrscheinlichkeit berechnet, dass das jeweilige Wort für den Unterschied verantwortlich ist. Der Anteil aller 500 Wörter am Gesamtunterschied zwischen zwei Texten addiert sich dabei jeweils auf 1. Aus diesen insgesamt 3.655.000 Wahrscheinlichkeitswerten ($86 \times 85 \times 500$) lässt sich dann die Entropie für jede der 7310 Textbeziehungen ermitteln. Abb. 13 zeigt sie in einem *Scatterplot* in Relation zum Delta-Wert, sodass wir zwei Informationen gleichzeitig sehen können: Wie groß der Unterschied (Delta-Wert) ist und wie sehr er sich auf einzelne Wörter (Entropie-Wert) konzentriert. Die Entropie ist klein, wenn nur wenige Wörter für den Unterschied verantwortlich sind; sie ist groß, wenn viele Wörter einen Beitrag leisten.

Auffällig ist, dass sich unsere vier unähnlichen Texte (Scheffel, Wolf, Wallner und Riehl) in Beziehung zueinander und zu Goldammers *Eine Hochzeitsnacht* vom Rest des Korpus dadurch unterscheiden, dass sie bei einem hohen Delta-Wert einen niedrigen Entropie-Wert aufweisen. Der große Unterschied wird hier also in der Tat durch wenige Wörter verursacht, sodass eine Distanzberechnung über das maximal diskriminative Wort nicht abwegig erscheint. Allerdings hat diese Einsicht einen aller Wahrscheinlichkeit nach unglamourösen Grund, denn mit Ausnahme Riehls sind die unähnlichen auch die kürzesten Texte in unserem Korpus. Wie auch in anderen Zusammenhängen können kurze Texte leicht zu Scheineinsichten führen, wenn die nicht ganz so häufigen Wörter in ihnen unterrepräsentiert sind. Ein Versuch mit einem Subset nur der mittellangen Texte hat freilich, was die Entropiespreizung angeht, kein signifikant anderes Ergebnis erbracht.⁴⁶ Wir werden diese Idee weiterverfolgen.

⁴⁵Die Idee dazu stammt von meinem Projektpartner Ulrik Brandes (ETH Zürich).

⁴⁶Zur Einordnung der Werte auf der Y-Achse: Das Entropie-Minimum, bei dem ein einziges Wort allen Unterschied ausmacht, liegt bei 0, das Maximum, bei dem sich der Unterschied gleichmäßig auf alle 500 Wörter verteilt, bei 6,2.

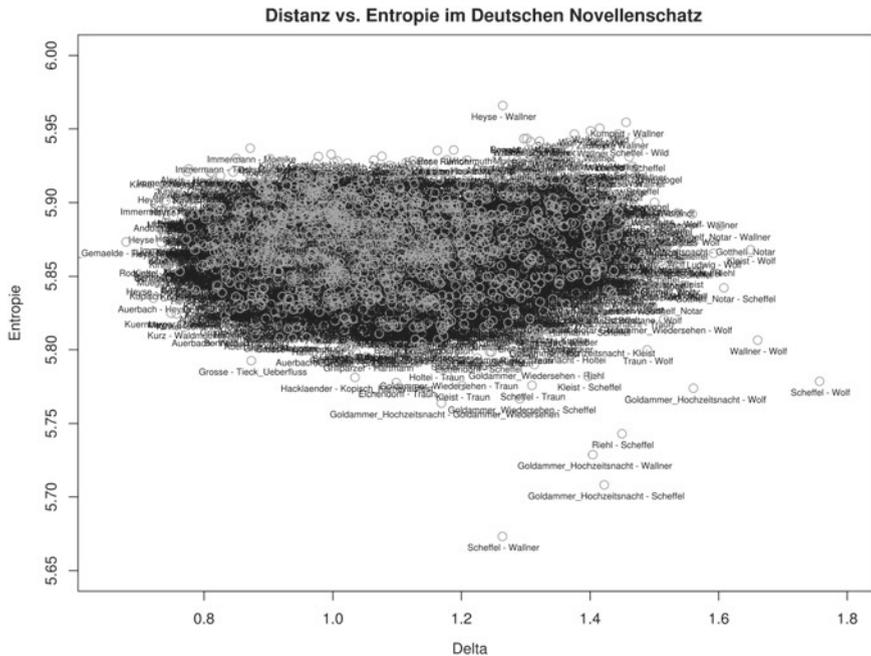


Abb. 13 Entropie/Delta-Plot für 500 MFW des Novellenschatzes³⁷

Literatur⁴⁷

- Bonter, Urszula, *Paul Heyse. Hofdichter und Publikumsschriftsteller*, in: Krzysztof Ruchniewicz/ Marek Zybura (Hg.), *Die höchste Ehrung, die einem Schriftsteller zuteil werden kann. Deutschsprachige Nobelpreisträger für Literatur*, Dresden 2007, 61–88.
- Brandes, Ulrik/Robins, Garry/McCraine, Ann u.a., „What is Network Science?“, in: *Network Science* 1 (2013), 1–15.
- Brandes, Ulrik/Nick, Bobo/Lee, Conrad u.a., „Simmelian Backbones: Amplifying Hidden Homophily in Facebook Networks“, in: *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (2013), 525–532.
- Burrows, John F., „Delta“. A Measure of Stylistic Difference and a Guide to Likely Authorship“, in: *LLC* 17 (2002), 267–287.
- Eder, Maciej, „Visualization in stylometry. Cluster analysis using networks“, in: *Digital Scholarship in the Humanities* 32/1 (2017), 50–64.
- Evert, Stefan/Jannidis, Fotis/Pielström, Steffen u.a., „Towards a better understanding of Burrows’s Delta in literary authorship attribution“, in: *Proceedings of NAACL-HLT Fourth Workshop on Computational Linguistics for Literature* (2015), 79–88.

³⁸ Burrows-Delta, 500 MFW, 20 % Culling.

⁴⁷ Sämtliche digitalen Referenzen wurden letztmalig am 12.7.2022 eingesehen.

- Ejchenbaum, Boris, „Zur Theorie der formalen Methode“, in: Ders., *Aufsätze zur Theorie und Geschichte der Literatur*, Frankfurt a.M. 1965, 7–52.
- Groeben, Norbert, „Was kann/soll ‚Empirisierung in der Literaturwissenschaft‘ heißen?“, in: Philip Ajouri/Katja Mellmann/Christoph Rauen (Hg.), *Empirie in der Literaturwissenschaft*, Münster 2013, 47–74.
- Heyse, Paul/Kurz, Hermann (Hg.), *Der Deutsche Novellenschatz*, 24 Bde., München/Leipzig 1871–1876.
- Jannidis, Fotis/Pielström, Steffen/Schöch, Christof u.a., „Improving Burrows‘ Delta – An empirical evaluation of distance measures“, in: *Digital Humanities Conference Sydney* (2015), https://www.researchgate.net/publication/280086768_Improving_Burrows%27_Delta_-_An_empirical_evaluation_of_text_distance_measures.
- Jeziorkowski, Klaus, „Der Virtuose des Durchschnitts. Der Salonautor in der deutschen Literatur des 19. Jahrhunderts, dargestellt am Beispiel Paul Heyse“, in: Ders., *Eine Iphigenie rauchend. Aufsätze und Feuilletons zur deutschen Tradition*, Frankfurt a.M. 1987, 114–129.
- Jockers, Matthew, *Macroanalysis. Digital Methods and Literary History*, Chicago 2013.
- Pennebaker, James W., *The Secret Life of Pronouns. What our Words Say About Us*, New York 2011.
- Walkhoff, Monika, *Der Briefwechsel zwischen Paul Heyse und Hermann Kurz in den Jahren 1869–1873 aus Anlass der Herausgabe des Deutschen Novellenschatzes*, München 1967.
- Weitin, Thomas, „Selektion und Distinktion. Paul Heyse's und Hermann Kurz' *Deutscher Novellenschatz* als Archiv, Literaturgeschichte und Korpus“, in: Daniela Gretz/Nicolas Pethes (Hg.), *Archiv/Fiktionen. Verfahren des Archivierens in Literatur und Kultur des langen 19. Jahrhunderts*, Freiburg i.Br. 2016, 385–408.
- Weitin, Thomas, „Literarische Heuristiken. Die Novelle des Realismus“, in: Albrecht Koschorke (Hg.), *Komplexität und Einfachheit. DFG-Symposium 2015*, Stuttgart 2017, 422–441.
- Weitin, Thomas, „Average and Distinction. The *Deutsche Novellenschatz* between Literary History and Corpus Analysis“, in: *LitLab Pamphlet* 6 (März 2018), <http://www.digitalhumanitiescooperation.de/pamphlete/>.

Bibliographische Ergänzungen 2021

Empirie und Statistik in der Digitalen Literaturwissenschaft

- Algee-Hewitt, Mark/Allison, Sarah/Gemma, Marissa u.a., „Canon/Archive. Large-scale Dynamics in the Literary Field“, in: *Pamphlets of the Stanford Literary Lab* 11 (2016).
- Bode, Katherine, *A world of fiction. Digital collections and the future of literary history*, Ann Arbor 2018.
- Herrmann, J. Berenike/Lauer, Gerhard, „Korpusliteraturwissenschaft. Zur Konzeption und Praxis am Beispiel eines Korpus zur literarischen Moderne“, in: *Osnabrücker Beiträge zur Sprachtheorie* 92 (2018), 127–156.
- Kuhn, Jonas, „Computerlinguistische Textanalyse in der Literaturwissenschaft? Oder: ‘The Importance of Being Earnest’ bei quantitativen Untersuchungen“, in: Toni Bernhart/Marcus Willand/Sandra Richter u.a. (Hg.), *Quantitative Ansätze in den Literatur- und Geisteswissenschaften*, Berlin/Boston 2018, 11–44.
- Underwood, Ted: *Distant horizons: Digital Evidence and Literary Change*. Chicago 2019.

Stilometrie und hochdiskriminative Wörter

- Chaudhuri, Prमित/Dasgupta, Tathagata/Dexter, Joseph u.a., „A small set of stylometric features differentiates Latin prose and verse“, in: *Digital Scholarship in the Humanities* 34/4 (2019), 716–729.
- Evert, Stefan/Proisl, Thomas/Jannidis, Fotis u.a., „Understanding and explaining Delta measures for authorship attribution“, in: *Digital Scholarship in the Humanities* 32 (2017), ii4–ii16.
- Hoover, David L., „The microanalysis of style variation“, in: *Digital Scholarship in the Humanities* 32 (2017), ii17–ii30.

- Päpcke, Simon/Weitin, Thomas/Herget, Katharina u.a., “Stylometric Similarity in Literary Corpora. Non-Authorship Clustering and ‘Deutscher Novellenschatz’”, in: *Digital Scholarship in the Humanities* [forthcoming].
- Schöch, Christof, „Zeta für die kontrastive Analyse literarischer Texte Theorie, Implementierung, Fallstudie”, in: Toni Bernhart/Marcus Willand/Sandra Richter u.a. (Hg.), *Quantitative Ansätze in den Literatur- und Geisteswissenschaften*, Berlin/Boston 2018, 77–94.
- Weidman, Sean G./O’Sullivan, James, “The limits of distinctive words. Re-evaluating literature’s gender marker debate”, in: *Digital Scholarship in the Humanities* 33/2 (2018), 374–390.
- Weitin, Thomas, *Digitale Literaturgeschichte. Eine Versuchsreihe mit 7 Experimenten*. Berlin/Heidelberg (2021), <https://doi.org/10.1007/978-3-662-63663-3>.

Entropie

- Chen, Ruina/Liu, Haitao/Altmann, Gabriel, “Entropy in Different Text Types”, in: *Digital Scholarship in the Humanities* 32 (2017), 528–542.
- Estevez-Rams, Ernesto/Mesa-Rodriguez, Ania/Estevez-Moya, Daniel, “Complexity-entropy analysis at different levels of organisation in written language”, in: *PLOS ONE* 14/5 (2019), 1–16.

Online-Ressourcen

Visione, www.visione.info.

Weitin, Thomas, *Volldigitalisiertes XML-Korpus: Der Deutsche Novellenschatz*, hg. von Paul Heyse/Hermann Kurz, 24 Bände. 1871–1876, Darmstadt/Konstanz (2016), <http://www.deutschestextarchiv.de/doku/textquellen#novellenschatz>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Operationalisierung der Metapher zur quantifizierenden Untersuchung deutschsprachiger literarischer Texte im Übergang vom Realismus zur Moderne

J. Berenike Herrmann

1 Einleitung

Der vorliegende Beitrag möchte eine ganz spezifische Antwort auf die Frage ‚Wozu eine Digitale Literaturwissenschaft?‘ geben. Ausgehend von dem Desiderat, Literarizität in einem digitalen, datengetriebenen Forschungsparadigma beschreiben, modellieren und mittels Verfahren der Künstlichen Intelligenz eventuell auch vorhersagen zu können, widme ich mich dem Phänomen metaphorischen Sprachgebrauchs in literarischen Texten. Es handelt sich um eine induktive Studie an einem kleineren Korpus deutschsprachiger Eingangspassagen narrativer fiktionaler Texte im Übergang vom Realismus zur Moderne. Ziel ist die datenbasierte Beschreibung der Bedingungen von Bedeutungskonstitution durch metaphorisch gebrauchte Lexeme. Die Metapher wird zu diesem Zweck als Kontrast und Ähnlichkeit zwischen zwei synchronen Wortbedeutungen operationalisiert, um durch den kontrollierten Einsatz eines korpusbasierten Lexikons (*Duden*) möglichst annotatorenunabhängig identifiziert¹ werden zu können. Trotz der manuellen Verfahrensweise handelt es sich hierbei also nicht um ein *Close Reading* im herkömmlichen Sinne.² Vielmehr wird hier erstens eine *händische Form des Distant Reading* in Anschlag gebracht – eine

¹J. Berenike Herrmann/Karola Woll/Aletta G. Dorst, „Linguistic Metaphor Identification in German“, in: Susan Nacey/Aletta G. Dorst/Tina Krennmayr u. a. (Hg.), *MIPVU in Multiple Languages*, Amsterdam 2019, 113–136.

²J. Berenike Herrmann, „Data-Driven Literary Studies“ (in Vorbereitung).

J. B. Herrmann (✉)
Universität Bielefeld, Bielefeld, Deutschland
E-Mail: berenike.herrmann@uni-bielefeld.de

algorithmische, stark kontrollierte Lesart, die die Interpretation aufschiebt und die potenziell zu automatisieren ist.³ An die Phase der Identifikation schließt sich zweitens die statistische Analyse an, die mögliche überzufällige Zusammenhänge mit weiteren formalisierten Textmerkmalen (Wortart und morphosyntaktische Komplexität) überprüft. Erst im dritten Schritt, und hier ist der Unterschied zum hermeneutischen *Close Reading* am deutlichsten, folgt die Interpretation, nämlich die der aggregierten Daten. Zugrunde liegt dem Verfahren die Annahme, dass Bedeutungskonstitution durch literarische Artefakte regelhaft ist⁴ und dass Regelmäßigkeit eben auch für Metaphern gilt.⁵

Literarizität wird im Folgenden anhand metaphorischen Sprachgebrauchs auf eine replizierbare und vom Beobachter möglichst unabhängige Art und Weise operationalisiert. Dabei soll die Metapher jedoch gerade nicht auf eine Abweichungsästhetik im Sinne Mukařovskýs und Šklovskijs beschränkt bleiben: Im Vordergrund steht nicht die Beschreibung der Konstitutionsregeln des auffälligen, abweichenden oder kreativen metaphorischen Ausdrucks – sondern eine Bestandsaufnahme des gesamten Spektrums zwischen konventionalisiert-lexikalisiertem und abweichend-neologischem Metapherngebrauch. Ausgehend von der Grundidee der konzeptuellen Metapherntheorie nehme ich an, dass die systematischen und regelhaften figurativen Bedeutungsrelationen des Sprachsystems⁶ auch für den literarischen Diskurs gelten.⁷ Dabei ist es eine empirische Frage, wie genau die metaphorischen konzeptuellen Strukturen und Prozesse sowie ihre sprachlichen Indikatoren im literarischen Diskurs aussehen.⁸ Obwohl die Metaphernforschung traditionell die Unterschiede zwischen Metaphern in literarischen und nicht-literarischen Werken hervorgehoben hat, gibt es eben auch eine Tradition der Beschreibung der Kontinuität zwischen beiden Bereichen,⁹ an die hier radikalisiert angeknüpft werden soll.

Mit der Kognitiven Theorie der Metapher (KTM) sind systematische und kohärente figurative Relationen beschreibbar geworden, die lokal auf intra-textueller Ebene, aber auch auf Diskurs- und Sprachsystemebene konstatiert

³Vgl. Petra Gehring/Iryna Gurevych, „Suchen als Methode? Zu einigen Problemen digitaler Metapherndetektion“, in: *Journal Phänomenologie* 41 (2014), 99–109.

⁴Vgl. Fotis Jannidis/Gerhard Lauer/Matías Martínez u. a., „Der Bedeutungsbegriff in der Literaturwissenschaft. Eine historische und systematische Skizze“, in: Dies. (Hg.), *Regeln der Bedeutung. Zur Theorie der Bedeutung literarischer Texte*, Berlin 2003, 3–30.

⁵Vgl. Rüdiger Zymner, „Uneigentliche Bedeutung“, in: Jannidis/Lauer/Martínez u. a. (Anm. 4), 128–168.

⁶George Lakoff/Mark Johnson, *Metaphors we Live by*, Chicago 1980.

⁷George Lakoff/Mark Turner, *More Than Cool Reason. A Field Guide to Poetic Metaphor*, Chicago 1989.

⁸Vgl. Elena Semino/Gerard J. Steen, „Metaphor in Literature“, in: Raymond W. Gibbs Jr. (Hg.), *Cambridge Handbook of Metaphor and Thought*, Cambridge, UK, 2008, 232–246; vgl. auch Gerard J. Steen/Aletta G. Dorst/J. Berenike Herrmann, u. a., „Metaphor in Usage“, in: *Cognitive Linguistics* 21/4 (2010), 765–796.

⁹Vgl. Semino/Steen (Anm. 7).

werden können. Ausgehend von der Ubiquität von konzeptuellen Metaphern wie DAS LEBEN IST WIE EINE REISE schlagen Lakoff und Turner vor, dass in literarischen Kontexten spezifische Verfahren angewandt werden, nämlich Erweiterung, Elaboration, Hinterfragen und (Neu-)Zusammensetzung von konventionellen Metaphern. Diese Verfahren führen die Leser literarischer Texte „beyond the bounds of ordinary modes of thought“ und „beyond the automatic and unconscious everyday use of metaphor“.¹⁰ Damit konzentrieren sich allerdings auch Lakoff und Turner auf die ‚Andersartigkeit‘ einer *poetischen* literarischen Metapher und knüpfen so an die abweichungsästhetische Tradition des russischen Formalismus an (allerdings ohne expliziten Verweis auf diese).

Wenn die konzeptuellen und sprachlichen Metaphern analytisch deutlicher getrennt werden als bei Lakoff und Kollegen, wird das gesamte Spektrum der sprachlichen Metaphernverwendung beschreibbar.¹¹ Verschiedene textorientierte Ansätze auch innerhalb des kognitiven Paradigmas haben Muster literarischer Metaphern auf der textuellen Ebene herausgearbeitet, etwa in Interaktion mit anderen Tropen wie der Metonymie bei Personifikationen¹² oder in lexiko-grammatikalischen Strukturen, wobei besonders Wortarten¹³ und häufige lexikalische Verbindungen¹⁴ sowie anaphorische Bezüge¹⁵ thematisiert werden.

Im Folgenden handelt es sich um eine Überprüfung der metaphorologischen Annahmen des russischen Formalismus unter gleichzeitiger Umkehrung seiner generellen Blickrichtung. Ausgehend von der Metapher als kennzeichnendem Merkmal für Literarizität¹⁶ geht es gerade darum, im literarischen Diskurs auch diejenigen Aspekte von metaphorischer Sprache aufzuzeigen, die auf den ersten Blick *nicht* von einer angenommenen Alltagssprache abweichen. Die Operationalisierung weist gleichermaßen kreative und konventionelle Metaphern aus – denn nur eine Beschreibung, die die Verteilung aller vorkommenden Elemente berücksichtigt, ist vollständig und dadurch eventuell sogar in der Lage, bisher übersehene Muster zu erkennen.

¹⁰Lakoff/Turner (Anm. 6), 72.

¹¹Semino/Steen (Anm. 7); vgl. Steen, Gerard J., *Finding Metaphor in Grammar and Usage*, Amsterdam/Philadelphia 2007.

¹²Aletta G. Dorst, „Personification in Discourse. Linguistic Forms, Conceptual Structures and Communicative Functions“, in: *Language and Literature* 20/2 (2011), 113–135.

¹³Aletta G. Dorst, *Metaphor in Fiction. Language, Thought and Communication*, Osterwijk 2011; Andrew Goatly, *The Language of Metaphors*, London/New York 1997; vgl. die nicht-kognitive Arbeit von Christine Brooke-Rose, *A Grammar of Metaphor*, London 1958.

¹⁴Alice Deignan, *Metaphor and Corpus Linguistics*, Amsterdam/Philadelphia 2005.

¹⁵Helge Skirl, „Metaphorical Anaphors“, in: Monika Schwarz-Friesel/Manfred Consten/Mareile Knees (Hg.), *Anaphors in Text. Cognitive, Formal and Applied Approaches to Anaphoric Reference*, Amsterdam 2007, 103–120.

¹⁶Z.B. Roman Jakobson, „Zwei Seiten der Sprache und zwei Typen aphatischer Störungen“, in: Ders., *Aufsätze zur Linguistik und Poetik*, München 1979, 135.

Metaphorizität wird zu weiteren Merkmalen der Sprachoberfläche in Beziehung gesetzt, nämlich zu Wortarten und dem *Flesch-Index* als standardisiertem Maß von (morpho-)syntaktischer Komplexität,¹⁷ neben einer groben Einordnung der Erzählperspektive (,vornehmlich erste Person‘ vs. ,vornehmlich dritte Person‘). In der Diskussion wird die statistische Analyse auf Ergebnisse der literaturwissenschaftlichen Forschung zu Erzähltexten der Jahre 1880–1930 bezogen. Mein Vorschlag betrifft so neben den konzeptionellen und technologischen Rahmenbedingungen digitaler Metaphernanalyse in kleineren Textumfängen auch literaturhistorische Fragestellungen: Wie verändert sich metaphorischer Sprachgebrauch in Erzählanfängen zwischen 1880 und 1930 in Korrelation mit weiteren quantifizierbaren Stilmerkmalen?¹⁸

Mit einer literaturwissenschaftlichen Anwendung der korpusbasierten Registervariationsstudien Bibers lege ich zugrunde, dass formalisierbare sprachliche Merkmale mit bestimmten Sprachregistern, Texttypen und Gattungen interagieren.¹⁹ Grundannahme ist, dass häufiges gemeinsames Auftreten von formal beschreibbaren Merkmalen eine gemeinsame kommunikative Funktion anzeigt. Für eine Stichprobe von literarischen Texten aus unterschiedlichen Subgenres des *Lancaster-Oslo/Bergen Corpus* konnte Biber etwa eine signifikante Kookkurrenz von „frequent past-tense forms, third-person personal pronouns, and perfect-aspect verbs“ sowie „public verbs, synthetic negation, present-participial clauses“ zusammen mit der besonders niedrigen Verwendung von „present-tense verbs and attributive adjectives“ zeigen.²⁰ Kein anderes Genre zeigte diese Zusammensetzung. Die Funktion dieses Ensembles von Merkmalen sieht Biber u. a. darin, vergangene Ereignisse und ihre andauernde Wirkung zu beschreiben (Vergangenheitsformen der Verben), Rede und Gedanken wiederzugeben (*Public Verbs*), die Figurenreferenz herzustellen (*Pronomina*) sowie anschauliche Beschreibungen zu liefern (*Present Participles*). Dorst verwandte ein ähnliches Verfahren zur Metaphernanalyse des literarischen Registers des *VU Amsterdam Metaphor Corpus (VUAMC)*, das im Registervergleich eine nur durchschnittliche Metaphernhäufigkeit auswies, darunter v. a. konventionelle Metaphern mit häufigen (verbalen und adjektivischen) Personifikationen.²¹ Eine signifikant häufige Verwendung von figurativen Vergleichen

¹⁷ Alan Bailin/Ann Grafstein, *Readability. Text and Context*, New York 2016.

¹⁸ J. Berenike Herrmann/Karina van Dalen-Oskam/Christof Schöch, „Revisiting Style, a Key Concept in Literary Studies“, in: *Journal of Literary Theory* 9/1 (2015), 25–52.

¹⁹ Vgl. Douglas Biber, *Variation Across Speech and Writing*, Cambridge, UK, 1988; Douglas Biber, „A Typology of English Texts“, in: *Linguistics* 27 (1989), 3–43; Douglas Biber/Edward Finegan, „Drift and the Evolution of English Style. A History of Three Genres“, in: *Language* 65 (1989), 487–517; Douglas Biber/Edward Finegan, „Multi-Dimensional Analyses of Authors’ Styles. Some Case Studies from the Eighteenth Century“, in: Don Ross/Dan Brink (Hg.), *Research in Humanities Computing*, Oxford 1994, 3–17; Susan Conrad/Douglas Biber (Hg.), *Variation in English. Multi-Dimensional Studies*, London 2001; Jesse Egbert, „Style in Nineteenth Century Fiction. A Multi-Dimensional Analysis“, in: *Scientific Study of Literature* 2/2 (2012), 167–198.

²⁰ Biber 1988 (Anm. 18), 102 u. 108 f.

²¹ Dorst (Anm. 12); vgl. Gerard J. Steen/Aletta G. Dorst/J. Berenike Herrmann u. a., „VU Amsterdam Metaphor Corpus“, in: *Oxford Text Archives*, Oxford 2010.

fiel auf, und damit ein lexikalisch markierter, möglicherweise augenfälliger, ‚literarischer‘ Metapherentypus.

Wie deutlich wird, erlaubt diese Form der datengetriebenen digitalen Literaturwissenschaft einen vergleichenden Ansatz, der nach Bedarf bestimmte Variablen fokussieren kann. Im Sinne der Russischen Formalisten können Generalisierungen über ‚Literatur‘ als sprachliches Register vorgenommen werden, ebenso wie über Gattungen (‚Romane‘, ‚Erzählungen‘) und Textpositionen (‚Exposition‘) – und auch über literaturhistorische Periodisierungen wie ‚Realismus‘ und ‚Moderne‘. Ebenso kann mit der Deviationsstilistik nach Voßler und Spitzer auf die Individualität einzelner literarische Texte (und Autoren bzw. Autorinnen) eingegangen werden, indem bestimmte textuelle Merkmale, in unserem Fall also der Gebrauch metaphorischer Ausdrücke, einen Stil anzeigen. Hier sind nicht unbedingt augenfällige Metaphern kennzeichnend, sondern ggf. unauffällige, aber quantitativ aussagekräftige Verteilungsmuster. Erst vor dem Hintergrund registerübergreifender Beschreibungen von Regeln und Strukturen der metaphorischen Bedeutungskonstitution können solche aufleuchten und spezifisch literarische Muster (bzw. Autor-, Gattungs-, Textpositions- oder Epochenstil) anzeigen.

Im empirischen Teil dieses Aufsatzes, der das Erzählanfangskorpus (EAK)²² als Datensatz zugrunde legt, werden einige Fragen zu metaphorischem Sprachgebrauch in fiktionalen Erzählanfängen des Zeitraums 1880–1926 exemplarisch ausgelotet. Dieses *Distant Reading* ist nicht allein durch die (digitale) Technik – oder den reinen Umfang der Daten definiert –, sondern durch das Streben nach Validierung durch die wissenschaftliche Methode.²³ Ähnlich zeigt auch Underwood den Bezug des *Distant Reading* zum quantitativen sozialwissenschaftlichen Forschungsparadigma auf.²⁴ Sein Ansatz fordert allerdings als hinreichende Bedingungen nicht nur die Qualitätskriterien Wiederholbarkeit und Beobachterunabhängigkeit, sondern zudem ein explanatives Vorgehen: „using hypotheses and samples (of texts or other social evidence) that are defined before the writer settles on a conclusion“.²⁵ Dies schließt explorative Ansätze, wie sie unter unterschiedlichen Vorzeichen etwa innerhalb der Autorschaftserkennung²⁶

²²J. Berenike Herrmann, *Erzählanfangskorpus (EAK)*. Elektronische Ressource, <https://github.com/jberenike/EAK> (letzter Aufruf 1.9.2017).

²³J. Berenike Herrmann, „In a Test Bed with Kafka. Introducing a Mixed-Method Approach to Digital Stylistics“, in: Joris J. van Zundert/Sally Chambers/Marijn Koolen/Mike Kestemont/Catherine Jones (Hg.), *DHQ: Digital Humanities Quarterly* 11/4 (2017), <http://digitalhumanities.org/dhq/vol/11/4/000341/000341.html> (letzter Aufruf 1.9.2017).

²⁴Ted Underwood, „A Genealogy of Distant Reading“, in: *Digital Humanities Quarterly* 11/2 (2017).

²⁵Ebd., 5.

²⁶Moshe Koppel/Jonathan Schler/Shlomo Argamon, „Computational Methods in Authorship Attribution“, in: *Journal of the American Society for Information Science and Technology* 60/1 (2008), 9–26.

und in der Korpusstilistik²⁷ verwendet werden, meines Erachtens zu Unrecht aus.²⁸ *Distant Reading* wird im Folgenden, wie oben beschrieben, als *empirische* Operation aufgegriffen, die die regelgeleitete Identifikation der Merkmale und ihre statistische Auswertung der Interpretation vorordnet.

2 Theoretische Modellierung: Die literarische kognitive Metapher

Die Metapher ist aus Theorien von Literarizität nicht wegzudenken. Als Mittel des Redeschmucks (*ornatus*), aber noch allgemeiner als ‚bildhafte Sprache‘ steht sie im Zentrum von Literaturtheorien, die literarische Kunst als gelungene Bildhaftigkeit auffassen, sie verkörpert wie kein zweites Verfahren ein vielfach reproduziertes Verständnis von *ut pictura poesis* (vgl. Horaz‘ *Ars Poetica*), das eine „sinnliche[...] Vergegenwärtigung“²⁹ des Beschriebenen leisten kann. Diese Funktion gilt seit der Antike als eine der wichtigsten der (poetischen) Metaphern, wie etwa Lausberg mit Ciceros *De Oratore* und Aristoteles‘ *Rhetorik* und *Poetik* zeigt.³⁰ Gerade wegen dieses sinnlichen Vergegenwärtigungscharakters spielt die Metapher interessanterweise gleichermaßen in Mimesis- und in Abweichungstheorien der Literatur eine zentrale Rolle. So kommt beispielsweise Šklovskij, der sich in *Kunst als Verfahren* eigentlich gerade von der Definition ‚Kunst ist Denken in Bildern‘ abgrenzt, um die Metapher nicht herum, ist sie doch ein hervorragendes Mittel der Verfremdung.³¹ Freilich unter den Vorzeichen einer

²⁷ Michaela Mahlberg, „Corpus Stylistics. Bridging the Gap between Linguistic and Literary Studies“, in: Michael Hoey/Michaela Mahlberg/Michael Stubbs u. a. (Hg.), *Texts, Discourse and Corpora*, London 2007, 219–246.

²⁸ Bode kritisiert am *Distant-Reading*-Paradigma besonders eine reduktionistische und reifizierende Modellierung des Literatursystems durch scheinbar stabile Datenpunkte wie Erstpublikation und Autornationalität. Sie spricht damit wichtige Probleme der Validität von Operationalisierungen an. (Katherine Bode, „The Equivalence of ‚Close‘ and ‚Distant‘ Reading; or, Toward a New Object for Data-Rich Literary History“, in: *Modern Language Quarterly* 78/1 [2017], 77–106). *Distant Reading* ist jedoch eine quantitative Operation, die zwangsläufig Komplexität reduzieren muss. Wie überzeugend dies geschieht, hängt neben einem transparenten Umgang mit Forschungsdaten besonders von der Operationalisierung der verwendeten Konstrukte ab, seien dies ‚Werk‘, ‚Autornationalität‘ oder eben ‚Stilmerkmal‘. Zudem, und dies gilt für quantitative ebenso wie für hermeneutische Ansätze, gilt das Primat der Reflexion, in deren Rahmen die Bedingungen und Beschränkungen gewählter Vorgehensweisen ins Feld geführt sein wollen.

²⁹ Heinrich Lausberg, *Handbuch der literarischen Rhetorik. Eine Grundlegung der Literaturwissenschaft*, Stuttgart ³1990, 287.

³⁰ Vgl. auch Ekkehard Eggs, „Metapher“, in: Gert Ueding (Hg.), *Historisches Wörterbuch der Rhetorik*, Bd. 5, Tübingen 2000, 1109–1183.

³¹ Viktor Šklovskij, „Die Kunst als Verfahren“, in: Jurij Striedter (Hg.), *Russischer Formalismus. Texte zur allgemeinen Literaturtheorie und zur Theorie der Prosa*, München ⁵1994 (russ. 1916), 3–35.

Avantgarde-Ästhetik unterscheidet er so zwischen dem ‚alltäglichen Denken in Bildern‘, das eine „besondere Ökonomie der Geisteskräfte“³² erlaube, und dem ‚dichterischen Bild‘, das „in besonderen Verfahren hergestellt“ und „mit größtmöglicher Sicherheit als künstlerisch wahrgenommen“³³ wird. Aus Sicht des Russischen Formalismus ist nur die kreative Metapher als Verfremdungstechnik geeignet, „um das Empfinden des Lebens wiederherzustellen, um die Dinge zu fühlen, um den Stein steinern zu machen“.³⁴

Heute stehen die Allgegenwärtigkeit und der graduelle Charakter metaphorischer Sprache und bildhaften Denkens im Mittelpunkt einer interdisziplinären semiotisch ausgerichteten Metaphorologie,³⁵ die das Spektrum von der Katachrese über das ‚alltägliche Bild‘ bis hin zum ‚dichterischen Bild‘ auf mehrfachen Ebenen modelliert. Grundannahme ist immer die des Potenzials einer ‚sinnlichen Vergegenwärtigung‘ – sprachlich indizierte konkrete Erfahrungsbereiche, insbesondere die der direkten sinnlichen Wahrnehmung, gelten als Grundlage für Übertragungsleistungen auf psychologischer und semiotischer Ebene.

In der vorliegenden Studie arbeite ich mit einer Operationalisierung des Konzepts der Kognitiven Theorie der Metapher (KTM), das die Metapher prozesshaft fasst („understanding and experiencing one kind of thing in terms of another“³⁶). Die Produkte dieser Prozesse sind Denk- bzw. Erfahrungsfiguren, die als konzeptuelle *Mappings*, Abbildungen einer Quell- auf eine Zieldomäne, aus „sets of conceptual correspondences“³⁷ bestehen. Lakoff erläutert die Idee der Korrespondenzsätze zwischen Quell- und Zieldomäne am Beispiel des LOVE-AS-JOURNEY³⁸-*Mappings*:

„The lovers correspond to travelers. The love relationship corresponds to the vehicle.
The lovers’ common goals correspond to their common destinations on the journey.
Difficulties in the relationship correspond to impediments to travel.“³⁹

Metaphern gelten auch in ihren kreativen literarischen Manifestationen als Ausdruck von konventioneller Sprache und Kognition.⁴⁰ Quelldomänen der metaphorischen *Mappings* sind in der Regel basale Erfahrungsbereiche des

³² Ebd., 3.

³³ Ebd., 7.

³⁴ Ebd., 15.

³⁵ Winfried Nöth, *Handbuch der Semiotik*, Stuttgart/Weimar 2000.

³⁶ Lakoff/Johnson (Anm. 5), 5 [Kursivierungen im Original; J.B.H.].

³⁷ Lakoff, George, „The Contemporary Theory of Metaphor“, in: Andrew Ortony (Hg.), *Metaphor and Thought*, Cambridge, UK, ²1993, 202–251, hier: 207.

³⁸ Konzeptuelle Metaphern werden in der kognitionswissenschaftlichen Literatur gemeinhin in KAPITÄLCHEN notiert.

³⁹ Lakoff (Anm. 36).

⁴⁰ Lakoff/Johnson (Anm. 5).

Menschen: Dazu gehören das Hantieren mit Objekten, die Erfahrung des eigenen Körpers im Raum, aber auch das soziale und kulturelle Wissen (REISEN oder RESTAURANTBESUCHE).⁴¹ Die Zieldomänen weisen in der Regel einen höheren Abstraktheitsgrad auf (LOVE), können aber auch konkret sein, insbesondere im Fall der sogenannten *Image Mappings*, wo visuelle Eigenschaften eines Körpers auf einen anderen abgebildet werden (etwa ‚die Sonne ist eine Zitrone‘).⁴²

Die ursprünglichen Annahmen der konzeptuellen Metaphertheorie⁴³ sind in fast vier Jahrzehnten nuanciert worden, nicht zuletzt durch korpus- und psycholinguistische Studien.⁴⁴ Heute verständigt man sich auf eine *Potenzialität* von experientiell definierter Metaphorik, die auf einer symbolisch-strukturellen Beschreibungsebene dokumentiert wird und es zulässt, das Paradox ‚nicht-metaphorisch verarbeiteter Metaphern‘ zu modellieren:

„this symbolic description does not force researchers to conclude that all of these structures necessarily have to be actualized in cognitive processing by each and every individual. Nor can it be concluded that they therefore are not metaphorical if these structures do not trigger cross-domain mappings in processing.“⁴⁵

Das metaphorische Bedeutungspotenzial lexikalisierter Metaphern liegt in der kontextuellen *Remotivierbarkeit* und so im Aufrufen basaler Erfahrungsschemata. Beispiel (1), paraphrasierbar als ‚die Ehepartner trennten sich‘, stammt aus einem Wochenzeitschriftenartikel. Der Ausdruck indiziert eine konzeptuelle Metapher für den Erfahrungsbereich LEBEN bzw. LIEBE/PARTNERSCHAFT durch Wörter mit hochkonventioneller metaphorischer Bedeutung, die zudem in relativ fester phraseologischer Beziehung stehen:

⁴¹ Ebd.; Lakoff (Anm. 36).

⁴² Vgl. Lakoff (Anm. 36), 229.

⁴³ Lakoff und Johnson sind für die Unterspezifizität des epistemologischen Status der konzeptuellen Metapher kritisiert worden, verschmelzen sie doch die analytischen Ebenen von Denken und Sprache, was impliziert, dass auch Ausdrücke wie *in drei Jahren* zwangsläufig als kognitiver Vergleich zwischen ZEIT und PHYSISCHEM OBJEKT (BEHÄLTER) verarbeitet werden. Einen Überblick gibt Gerard J. Steen (Anm. 10); siehe auch Herrmann (Anm. 44).

⁴⁴ Vgl. Gibbs' Unterscheidung von vier Phasen des Echtzeit-Metaphernverstehens: (1) „comprehension“, (2) „recognition“, (3) „interpretation“ und (4) „appreciation“ (255 f.), wobei Phase (2) bis (4) auf eine unmittelbare Verarbeitung aufbauen und inkrementell elaborierter sind (Raymond W. Jr Gibbs, „Process and Products in Making Sense of Tropes“, in: Andrew Ortony (Hg.), *Metaphor and Thought*, Cambridge, UK, ²1993, 252–276).

⁴⁵ Gerard J. Steen, „The Paradox of Metaphor. Why we Need a Three-Dimensional Model for Metaphor“, in: *Metaphor and Symbol* 23/4 (2008), 213–41, hier: 218; vgl. J. Berenike Herrmann, *Metaphor in Academic Discourse. Linguistic Forms, Conceptual Structures, Communicative Functions and Cognitive Representations*, Utrecht 2013, 30.

1 Seit 1886 *gingen* die Ehepartner *getrennte Wege*.⁴⁶ (*Der Spiegel*, 14.04.1980, DWDS Kernkorpus 1980/sp19800414_8)⁴⁷

Um Wörter wie „gingen“, „getrennte“, „Wege“ als ‚potenziell metaphorisch verwendet‘ zu identifizieren, müssen zwei Kriterien erfüllt sein: erstens eine stabile Wortbedeutung, die auf einer basal(er)en Erfahrungsebene konkrete Vorgänge, belebte Entitäten und/oder Objekte referenzialisiert (z. B. „Weg“: ‚natürliche, durch Festtreten entstandene oder künstlich angelegte, nicht oder nur wenig dauerhaft befestigte, relativ schmale Bahn, besonders für Fußgänger‘); zweitens eine abweichende kontextuelle Bedeutung, die zur erstgenannten basaleren Bedeutung in einem klaren Kontrast steht, aber auch eine Ähnlichkeit aufweist (‚Lebensweg, Verlauf eines Lebens‘).⁴⁸

Obwohl die metaphorische Bedeutung von „Wege“ (1) im deutschen Sprachsystem lexikalisiert ist, kann der ‚Verlauf des Lebens‘ unter bestimmten diskursiven Bedingungen *wie* eine physisch wahrnehmbare Bewegung von Entitäten durch den Raum verstanden und evtl. sinnlich nachvollzogen werden. Das Hauptargument ist hier, dass eine konkret(er)e, oftmals körperorientierte und spezifischere Grundbedeutung eines Wortes im Wortschatz des untersuchten Sprachstandes vorhanden ist (operationalisiert durch die Auffindbarkeit in Wörterbüchern) und so im Diskursereignis zur potenziellen Remotivierung im Vergleich zur Verfügung steht.

Wie schon erwähnt, können trotz der Selbststilisierung der KTM als traditionsloser Neuerung klare Bezüge zu antiker Rhetorik einerseits und Russischem Formalismus andererseits aufgezeigt werden, und zwar bezüglich der Kriterien ‚Ähnlichkeit‘, ‚Konventionalität‘ und ‚Notwendigkeit‘. Die Ähnlichkeit zwischen Erfahrungsbereichen bzw. Wortbedeutungen, die auch die neuere empirische Metaphernforschung zentral stellt,⁴⁹ findet sich einschlägig z. B. bei Lausberg,⁵⁰ der die Metapher nach Quintilian als ‚*brevisitas*-Form des Vergleichs‘ fasst, wobei die Ähnlichkeit (*similitudo*) ein Abstufungsphänomen ist. Auch alltägliche Konventionalität von Metaphern kann bei Lausberg lokalisiert werden: ‚Metaphern stehen in Traditionen, die der jeweils okkasionellen Bedeutung *ein gewisses* (etwa durch die literarische Gattung bedingtes) *habituelles Gepräge*

⁴⁶Im Folgenden werden metaphorisch verwendete Lexeme durch Unterstreichungen hervorgehoben.

⁴⁷*Korpustreffer für „Weg“*, https://www.dwds.de/r?q=Weg;corpus=kern;format=full;date-start=1980;date-end=1999;genre=Belletristik;genre=Wissenschaft;genre=Zeitung;genre=Gebrauchsliteratur;p=4;sort=date_asc;limit=10 (letzter Aufruf 1.9.2017).

⁴⁸Vgl. das Lemma „Weg“ im DWDS, <https://www.dwds.de/wb/Weg> (letzter Aufruf 1.9.2017).

⁴⁹Vgl. z. B. Steen (Anm. 10); Gerard J. Steen/Aletta G. Dorst/J. Berenike Herrmann u. a., *A Method for Linguistic Metaphor Identification. From MIP to MIPVU*, Amsterdam/Philadelphia 2010, 37 f.

⁵⁰Lausberg (Anm. 28), 285.

geben“.⁵¹ Von hier ist es ein kleiner Schritt zur ‚notwendigen‘ Metapher, dem Hauptargument der KTM, das abstrakte Erfahrungsdomänen wie ZEIT durch ‚ontologische‘ Metaphern überhaupt erst konzeptualisier- und kommunizierbar macht, u. a. durch Präpositionen mit räumlicher Grundbedeutung („in drei Tagen“). Die *inopia*, der Mangel eines eigentlichen Ausdrucks, ist nach Lausberg eine „sehr häufige, weil für die Seinserkenntnis und -gliederung nützliche und notwendige semantische Erscheinung“.⁵² Es sei einleuchtend, „dass der geistige Realitäten bezeichnende Wortschatz katachrestischer Herkunft ist, *animus* („Wind“), *sapiens* („schmeckend“), *spiritus* („Hauch““).⁵³ Freilich spitzt die KTM solche Beobachtungen zu, wo sie metaphorische Übertragung als *Grundprinzip* von Kognition und Sprache modelliert. Dabei wird nach triftiger Kritik aus dem empirischen Lager⁵⁴ inzwischen mehr Spielraum beim Schluss von den sprachlichen Indizien auf die spezifischen konzeptuellen Strukturen zugelassen. Aus Beispiel (1) kann etwa das oben erwähnte LOVE-AS-JOURNEY *Mapping* abgeleitet werden, aber auch Alternativen, die sich im Grad der Generalisierung oder im Fokus unterscheiden:

ABSTRAKT IST WIE KONKRET

LEBEN IST WIE BEWEGUNG IM RAUM

LEBEN IST WIE EINE REISE

EIN SUBJEKT IST WIE EIN REISENDER

PARTNERSCHAFT IST WIE GEMEINSAME BEWEGUNG IM RAUM

Die literaturwissenschaftliche Anwendung der ‚klassischen‘ KTM schlägt vor, dass literarische wie nichtliterarische Metaphern basale somato-kognitive Schemata nutzen: „To the extent that a basic metaphor used in poetry is experientially grounded, it draws power from the fundamental nature of those experiences.“⁵⁵ Ähnlich wie Satz (1) indizieren auch der Vers aus dem österreichischen Barockgedicht (2) und der US-amerikanische Protestsong der 1960er (3) *Mappings* wie ABSTRAKT IST WIE KONKRET, LEBEN IST WIE BEWEGUNG IM RAUM und EIN SUBJEKT IST WIE EIN REISENDER:

- 2 Dann wollen mich die Wind auf andre Zufahrt dringen, / bring' an den Hafen mich/mein Gott/es ist genug! (Catharina Regina von Greiffenberg, *Auf meinen bestürmeten Lebens-Lauff*, 1662)
- 3 How many roads must a man walk down/Before you call him a man? (Bob Dylan, *Blowin' in the Wind*, 1962/1963)

⁵¹ Ebd., 288 [Kursivierungen im Original; J.B.H.].

⁵² Ebd., 290.

⁵³ Ebd.

⁵⁴ Z. B. John Vervaeke/John M. Kennedy, „Conceptual Metaphor and Abstract Thought“, in: *Metaphor and Symbol* 19/3 (2004), 213–232.

⁵⁵ Lakoff/Turner (Anm. 6), 84.

Hier werden *Wind*, *Zufahrt*, *Hafen* (2) und *roads* (3) ebenso wie *bring(en)*, *an* (2), *walk*, *down* (3), aber auch *wollen*, *dringen* und *auf* (2) metaphorisch verwendet. Für (2) beispielsweise können folgende Korrespondenzen abgeleitet werden:

EIN SUBJEKT IST WIE EIN REISENDER ZUR SEE
 EREIGNISSE SIND WIE WINDBÖEN
 EIN LEBENSZIEL IST WIE EIN HAFEN
 GOTT IST WIE EIN KAPITÄN

Literarische Metaphern unterscheiden sich also nicht grundlegend von Metaphern im Alltagsdiskurs, sondern sind deren Erweiterungen, Elaborationen, Infragestellungen und (Neu-)Zusammensetzung.⁵⁶ Im Unterschied zu (1) zeigen Beispiel (2) und (3) etwa eine gehäufte Rekurrenz von sprachlichen Referenzen auf die Quelldomänen (SCHIFFFAHRT, WANDERSCHAFT). Aber auch die genrebedingte Literarizitätserwartung der Leser ist ein ausschlaggebender Faktor, denn literarische Texte werden gemeinhin sorgfältiger und als bedeutungs-offener gelesen,⁵⁷ was Implikationen für die konkrete Verarbeitung auch von hochkonventionellen Metaphern als Präpositional- und Adverbialmetaphern hat. Dorsts *Fiction-Sample* des VUAMC⁵⁸ weist eine niedrigere Metaphernproportion als die akademischen und journalistischen Stichproben auf, doch treten metaphorische Adjektive, Nomen und Verben im Unterschied zu den anderen Registern mit einer besonders hohen lexikalischen Varianz auf. Die schon erwähnten Personifikationen und Vergleiche sind dabei Verfahren, die besonders häufig auftreten, nicht zuletzt zur Schaffung von „vivid images“.⁵⁹

Im Weiteren ist es mein Ziel, die Metapher zu operationalisieren, um Formen und Strukturen auf einer semiotischen Beschreibungsebene zu identifizieren.⁶⁰ Die Dimension der leserseitigen Verarbeitung literarischer metaphorischer Sprache spielt nach diesem Prinzip also keine Rolle. Sie kann in Interpretation der Ergebnisse spekulativ erörtert oder in Lesestudien gezielt untersucht werden.

⁵⁶ Ebd., 72; vgl. Semino/Steen (Anm. 7).

⁵⁷ Vgl. Gerard J. Steen, *Understanding Metaphor in Literature. An Empirical Approach*, London/ New York 1994.

⁵⁸ Dorst (Anm. 12), 186–203. Die Stichprobe besteht aus populären englischsprachigen Romanen des 20. Jahrhunderts, die Dorst in ein ‚realistisches‘ Genre einordnet.

⁵⁹ Dorst (Anm. 12), 268.

⁶⁰ Die im vorliegenden Aufsatz eingenommene Perspektive auf Metaphern als semiotisches Potenzial ist m. E. eine – besonders stark formalisierte – Ausprägung von Winkos „Verstehen 2“, nämlich dem „methodisch reflektierten Verstehen 2 der professionellen Teilnehmer des Literatursystems“ (Simone Winko, „Verstehen literarischer Texte versus literarisches Verstehen von Texten? Zur Relevanz kognitionspsychologischer Verstehensforschung für das hermeneutische Paradigma der Literaturwissenschaft“, in: *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* 69/1 (1995), 1–27, hier: 11), während Winkos „Verstehen 1 [...] das unmittelbare Verstehen“ (ebd., 7) betrifft, das in Bezug auf Metaphern durch Gibbs‘ (Anm. 43) Taxonomie von Verarbeitungsphasen noch weiter ausdifferenziert werden kann.

3 Operationalisierung: Annotationsprotokoll MIPVU

Die *Metaphor Identification Procedure VU University* (MIPVU)⁶¹ wurde für die präzise und systematische Identifikation metaphorischer Sprache auf der lexikalischen (Wort-)Ebene ursprünglich für das Englische und Niederländische entwickelt und für das Deutsche adaptiert.⁶² Sie ist vergleichbar mit einem sozialwissenschaftlichen Codierungsschema, das einen hohen Grad an Unabhängigkeit vom einzelnen Wissenschaftler und Replizierbarkeit ermöglicht. Ausschlaggebend sind die empirischen Qualitätskriterien der Intersubjektivität/Objektivität, Reliabilität und Validität. Die (statistische) Vergleichbarkeit über Analysen hinweg – und in einem gewissen Maße auch über unterschiedliche Sprachen – ist ein praktischer Vorteil für die strukturalistisch inspirierte Literaturanalyse. Weil MIPVU vor allem Spezifizierungen und Ergänzungen des ursprünglichen Protokolls MIP (*Metaphor Identification Procedure*)⁶³ enthält, werden im Folgenden die Guidelines für MIP vorgestellt und aus der Perspektive der deutschsprachigen MIPVU-Adaption kommentiert. Insbesondere müssen hierbei die Eignung der Analyseeinheit, die Adäquatheit der Beschreibung der (situationsspezifischen) Kontextbedeutung und der Belegbarkeit einer grundlegenden Bedeutung sowie der Grad der Distinktheit/Ähnlichkeit der beiden Bedeutungen berücksichtigt werden.⁶⁴

MIP – Metaphernidentifikationsprozedur (Pragglejaz Group)⁶⁵

1. Lies den gesamten Text-Diskurs, um dir ein allgemeines Verständnis seiner Bedeutung zu verschaffen.
2. Bestimme die lexikalischen Einheiten des textuellen Diskurses.
3. (a) Beurteile für jede lexikalische Einheit, welche Bedeutung sie im Kontext hat, d. h. inwiefern sie sich auf eine Entität, Relation oder ein Attribut in der vom Text evozierten Situation bezieht (kontextuelle Bedeutung). Achte dabei auch auf Vorhergehendes und Folgendes.
3. (b) Beurteile für jede lexikalische Einheit, ob es für sie in anderen Kontexten eine grundlegendere gegenwärtig-zeitnahe Bedeutung gibt als im gegenwärtigen Kontext. Für unsere Zwecke können als Grundbedeutungen solche Bedeutungen gelten, die:
 - konkreter sind [das Evozierte ist einfacher vorzustellen, zu sehen, hören, fühlen, riechen oder schmecken];
 - sich auf körperliche Tätigkeit beziehen;

⁶¹ Steen/Dorst/Herrmann u. a. 2010 (Anm. 48).

⁶² Herrmann/Woll/Dorst (Anm. 1).

⁶³ Pragglejaz Group, „MIP. A Method for Identifying Metaphorically Used Words in Discourse“, in: *Metaphor and Symbol* 22/1 (2007), 1–39.

⁶⁴ Vgl. Steen/Dorst/Herrmann u. a. 2010 (Anm. 48), 13 ff.

⁶⁵ Übersetzung durch die Autorin. Hier wird für die Übersetzung die Terminologie „-prozedur“ anstatt des eleganteren „-verfahren“ gewählt, da so die Nähe zum Akronym *MIP – Metaphor Identification Procedure* erhalten werden kann; siehe: Pragglejaz Group (Anm. 62), 3.

- präzise sind (im Gegensatz zu vage);
- historisch älter sind.

Grundbedeutungen sind nicht unbedingt die häufigsten Bedeutungen der lexikalischen Einheit.

3. (c) Wenn die lexikalische Einheit in anderen Kontexten als im gegebenen eine grundlegendere gegenwärtig-zeitnahe Bedeutung hat, muss entschieden werden, ob die kontextuelle Bedeutung im Kontrast zur Grundbedeutung steht, aber im Vergleich mit ihr verstanden werden kann.

Wenn ja, kann die lexikalische Einheit als metaphorisch markiert werden.

Um die Identifikation von Metaphern verlässlich durchzuführen, verwendet MIPVU als wichtigste externe Ressourcen korpusbasierte Diktionäre (für das Deutsche: *Duden* und *DWDS*), für die eine Reihe von Anwendungsregeln formuliert wurden. Zudem werden *Part-of-Speech-Tags* (POS) für die Identifikation von lexikalischen Einheiten und Wortarten berücksichtigt. MIPVU operiert wie MIP mit der Analyseeinheit ‚Wort‘ (MIP-Schritt 2), wobei Regeln die wörterbuchgestützte Identifikation stabiler lexikalischer Einheiten festlegen, z. B. bilden die abgetrennten Verbbestandteile von Partikel- und Präfixverben zusammen mit dem Verbstamm eine lexikalische Einheit, etwa „ablegen“ in ‚in Eurem Namen *lege* ich jetzt folgendes Versprechen *ab*‘.⁶⁶ Für jede identifizierte lexikalische Einheit unterscheidet MIPVU zwischen einer kontextuellen Bedeutung (MIP-Schritt 3a) und einem ‚more basic meaning‘ (MIP-Schritt 3b), einer grundlegenderen gegenwärtig-zeitnahen Bedeutung. Eine Reihe von Regeln definiert den Umgang mit Fällen, in denen die kontextuelle und/oder die basalere Bedeutung nicht oder nur schwer festgelegt werden kann,⁶⁷ u. a. steht für Grenzfälle ein besonderer Code zur Verfügung, das sogenannte WIDLII (‚When In Doubt Leave It In‘).⁶⁸ Kann die kontextuelle Bedeutung etabliert und mit einer identifizierbaren Grundbedeutung verglichen werden, ist aber auch deutlich von ihr abzugrenzen (MIP-Schritt 3c), handelt es sich um ein sogenanntes ‚metaphor-related word‘, ein MRW.⁶⁹

Für „Wege“ (Beispiel 1) kann mithilfe des Wörterbuchs die kontextuelle Bedeutung (‚hier gehen unsere Ansichten, Anschauungen so weit auseinander, dass unsere Zusammenarbeit o. ä. aufhört‘) klar von der Grundbedeutung (‚etwas, was wie eine Art Streifen – im Unterschied zur Straße oft nicht befestigt – durch ein Gebiet, Gelände führt und zum Begehen [und Befahren] dient‘) unterschieden

⁶⁶ Herrmann/Woll/Dorst (Anm. 1).

⁶⁷ Steen/Dorst/Herrmann u. a. (Anm. 48), 34 f.

⁶⁸ Ebd., 33 f.

⁶⁹ Das englische Akronym ‚MRW‘ wird aus Konsistenzgründen verwendet, mit MIP(VU) als Standard der Metapheridentifikation. Diese Terminologie schließt statt „metaphorically used“ oder „metaphor“ auch andere Formen wie Vergleiche, Analogien und Substitution ein – und zeigt nicht zuletzt die Potentialität der semiotischen Metaphorik gegen die empirische des Sprachverstehens an. Vgl. Steen/Dorst/Herrmann u. a. (Anm. 48).

werden.⁷⁰ Die beiden Bedeutungen sind deutlich verschieden, weisen aber eine Ähnlichkeitsbeziehung auf, sodass ein Vergleich konstituiert werden kann.⁷¹ Für das Deutsche MIPVU ist der *Duden* die Hauptreferenz, kann bei Unklarheiten aber durch das *DWDS* überstimmt werden.

MIPVU identifiziert keine ‚etymologischen‘ Metaphern, also Wörter, die zwar auf metaphorischem Wege entstanden sind, aber im gegenwärtigen Deutsch keine konkrete Grundbedeutung mehr haben. Ein Beispiel ist das umgangssprachliche Verb ‚zwiebeln‘, dessen Metaphorik etymologisch hergeleitet werden kann,⁷² das aber im gegenwärtigen Deutsch durch die Wörterbücher als monosem belegt wird (‚jemandem hartnäckig [mit etwas] zusetzen; schikanieren‘).⁷³ MIPVU identifiziert Metaphern also synchron und auf Wortebene – letzteres ist auch für fortgeführte Metaphern und Metaphernfelder ebenso wie für Analogien und Allegorien relevant: Die Bedeutung eines Wortes im Kontext wird jeweils mit seiner im Wörterbuch auffindbaren Grundbedeutung verglichen.⁷⁴

Für die Adaption von MIPVU mussten aufgrund der produktiven Wortbildungsmuster des Deutschen (Komposition, Konversion und Derivation) und der Wörterbuchstruktur wenige zusätzliche Regeln eingeführt werden.⁷⁵ Im Unterschied zum Englischen, welches durch besonders viele unabhängige Morpheme gekennzeichnet ist, hat das Deutsche mehr synthetische Elemente, auch weisen *Duden* und *DWDS* im Vergleich zum *Macmillan Dictionary* mehr Querverweise zwischen Lemmata über Wortarten hinweg auf. Da aber Wortartengrenzen für MIPVU ein ‚important factor of meaning differentiation‘⁷⁶ sind, ist hier ggf. eine etwas niedrigere Reliabilität der Methode für das Deutsche zu erwarten, da bei der

⁷⁰ <https://www.duden.de/rechtschreibung/Weg> (letzter Aufruf 20.5.2018). Siehe auch Herrmann/Woll/Dorst (Anm. 1) zu Regeln der Dudennutzung in Fällen wie ‚Weg‘.

⁷¹ Zusätzlich zur durch MIP vorgelegten Identifikation von ‚indirekten Metaphern‘, die durch indirekten Wortgebrauch Mappings wie LEBEN IST EINE REISE indizieren, enthält MIPVU Anweisungen zum Umgang mit ‚direkten Metaphern‘, die Mappings durch Lexis oder Kontext als Vergleich *direkt* anzeigen, und ‚impliziten Metaphern‘ mit Mappings durch lexikalische Substitution bzw. Ellipse. Vgl. Steen/Dorst/Herrmann u. a. (Anm. 48).

⁷² Pfeifers etymologisches Wörterbuch im *DWDS* führt unter dem Lemma ‚zwiebeln‘ ‚Vb. ‚jmdm. hart zusetzen, jmdn. Drillen‘ (17. Jh.), eigentl. wohl ‚jmdm. wie einer Zwiebel Schale für Schale abziehen‘, zuvor ‚mit Zwiebeln zubereiten‘ (16. Jh.)‘, <https://www.dwds.de/wb/zwiebeln> (letzter Aufruf 1.9.2017).

⁷³ <https://www.duden.de/rechtschreibung/zwiebeln> (letzter Aufruf 1.9.2017).

⁷⁴ MIPVU bleibt so nah an der Textoberfläche wie möglich, was für fiktionale/ästhetische Texte bedeutet, dass zunächst die Metaphorizität des Sprachgebrauchs auf der Ebene der Textwelt zu beurteilen ist. Bei deutlichen textuellen und/oder kontextuellen Hinweisen auf eine fortgeführte analogische oder allegorische Lesart, in der nicht-metaphorische Elemente der Textwelt gleichzeitig eine metaphorische Bedeutung auf einer weiteren Bedeutungsebene haben, kann statt des ‚Normalfalls‘ der indirekten Metapher eine ‚direkte Metapher‘ ausgezeichnet werden. Vgl. Steen/Dorst/Herrmann u. a. (Anm. 48), 14 f., 38 f. u. 57 f.

⁷⁵ Herrmann/Woll/Dorst (Anm. 1).

⁷⁶ Steen/Dorst/Herrmann u. a. (Anm. 48), 16 f.

Identifikation der Bedeutungen etwas mehr Interpretationsspielraum anzunehmen ist.⁷⁷ Zudem mussten wir mit Mehrwortausdrücken umgehen, die mit hoher Wahrscheinlichkeit lexikalisiert sind, für die die deutschen Diktionäre jedoch keine Lemmata anbieten (z. B. „ab und zu“). Diesen Schwierigkeiten zum Trotz zeigte ein Reliabilitätstest mit Kolleginnen aus der MIPVU-Arbeitsgruppe für eine erste deutschsprachige Version von MIPVU bereits vor der Diskussionsphase des Protokolls insgesamt eine moderate Übereinstimmung (Fleiss' *Kappa* von 0,73).⁷⁸ Das Fleiss' *Kappa* für den literarischen Text war niedriger (0,65) als das für den journalistischen Text (0,78).⁷⁹

4 Fallstudie EAK

4.1 Datengrundlage

Das EAK enthält insgesamt N=17,888 Wörter aus 35 Texten mit Erstpublikation 1880–1930, die Untergattungen wie Novelle und Roman abdecken, ebenso wie Unterhaltungs- und Avantgardeliteratur. Es enthält Texte von ‚modernen‘, aber auch als ‚realistisch‘ eingeordneten Autoren. Ausschlaggebend war i. d. R. das Erscheinungsdatum.⁸⁰ Als ausgewogene Stichprobe enthält das EAK Texte aus den deutschsprachigen Ländern, kanonische und auch populäre Texte sowie männliche

⁷⁷Vgl. Pasma für das Niederländische (Trijtje Pasma, *Metaphor and Register Variation. The Personalization of Dutch News Discourse*, Oosterwijk 2011).

⁷⁸Fleiss' *Kappa* ist ein Maß der Intercoder-Reliabilität, das für jede Instanz (hier: Lexem) die Übereinstimmung der Gesamtgruppe der Annotatoren erhebt. Nach McHugh (Marry L. McHugh, „Interrater Reliability. The Kappa Statistic“, in: *Biochemia Medica* [2012], 276–82) gilt für die *Kappa*-Werte: 0-0.20 *keine Übereinstimmung*; 0.21-0.39 *minimale Übereinstimmung*; 0.40-0.59 *schwache Übereinstimmung*; 0.60-0.79 *moderate Übereinstimmung*; 0.80-0.90 *starke Übereinstimmung*. Die Werte wurden mit der Software R (Version 3.3.3) errechnet, wobei die irr-package (Gamer/Lemon/Fellows/Singh) mit den Funktionen `kappam.fleiss(x)` (Fleiss' *Kappa*) and `agree(x)` (*Agreement*) genutzt wurde (Matthias Gamer/Jim Lemon/Ian Fellows u. a., „Irr. Various Coefficients of Interrater Reliability and Agreement“ [version 0.84, 2012], <https://CRAN.R-project.org/package=irr>).

⁷⁹Siehe auch Herrmann/Woll/Dorst (Anm. 1).

⁸⁰Hier liegt die Annahme eines „literarischen Strukturwandels vom späten Realismus zur frühen Moderne“ (Marianne Wunsch, *Realismus (1850–1890). Zugänge zu einer literarischen Epoche*, Kiel 2007) zugrunde, bei dem ab ca. 1880 in der „Spätphase des Realismus“ Texte auftreten, die Elemente moderner Erzählverfahren nutzen. Dies sind Texte, die insofern „an den Grenzen des *realistischen* Literatursystems sind, als sie mit Strukturen experimentieren, die in diesem Literatursystem eigentlich ausgeschlossen werden und normalerweise erst im folgenden literarischen System, der in den 1890ern sich allmählich konstituierenden *Frühen Moderne*, thematisch werden.“ (ebd., 297 [Kursivierungen im Original; J.B.H.]).

Tab. 1 Das EAK

TextID	AutorIn	Titel	Jahr	Wörter
1	Altenberg, Peter	Prödrömös	1906	486
2	Andreas-Salomé, Lou	Fenitschka. Eine Ausschweifung	1898	521
3	Bierbaum, Otto Julius	Stilpe	1897	501
4	Déry, Juliane	Selige Liebe	1896	502
5	Eschstruth, Nataly von	Katz' und Maus	1886	491
6	Fontane, Theodor	Schach von Wuthenow	1896	553
7	Fontane, Theodor	Effi Briest	1888	522
8	Fontane, Theodor	Der Stechlin	1899	526
9	Fontane, Theodor	Irrungen, Wirrungen	1883	508
10	Hagenauer, Arnold	Muspilli	1900	495
11	Heyking, Elisabeth von	Die Orgelpfeifen	1918	543
12	Kafka, Franz	Das Urteil	1913	563
13	Kafka, Franz	Die Verwandlung	1925	493
14	Kafka, Franz	Ein Hungerkünstler	1922	548
15	Kafka, Franz	Der Prozess	1925	517
16	Keller, Gottfried	Das Sinngedicht	1882	536
17	Keyserling, Eduard von	Beate und Mareile	1903	454
18	Kraft, Robert	Der Mediziner	1896	540
19	Kretzer, Max	Meister Timpe	1888	525
20	Laßwitz, Kurd	Auf zwei Planeten	1897	557
21	May, Karl	Durchs Wilde Kurdistan	1892	492
22	Meißner, Alfred	Die Prinzessin von Portugal	1882	521
23	Polenz, Wilhelm von	Der Büttnerbauer	1895	518
24	Preuschen, Hermione von	Yoshiwara	1920	482
25	Raabe, Wilhelm	Stopfkuchen	1896	520
26	Raabe, Wilhelm	Die Akten des Vogelsangs	1891	527
27	Reventlow, Fanny Gräfin zu	Herrn Dames Aufzeichnungen	1913	511
28	Schnitzler, Arthur	Fräulein Else	1924	466
29	Schnitzler, Arthur	Traumnovelle	1926	545
30	Schreiner, Olive	Peter Halket im Mashonalande	1898	544
31	Schubin, Ossip	Vollmondzauber	1899	518
32	Spyri, Johanna	Heidi's Lehr- und Wanderjahre	1880	513
33	Storm, Theodor	John Riew'	1887	520
34	Storm, Theodor	Ein Doppelgänger	1885	539
35	Sturza, Marie Tihanyi	Das Gelübde einer dreißigjährigen Frau	1905	461
Summe				18058
Mittelwert				515,94
(SD)				(26,55)

wie weibliche Autorschaft (siehe Tab. 1).⁸¹ Es wurden jeweils die ersten 600 Tokens (inkl. Satzzeichen) der Texte extrahiert, was etwa den 500 Wörtern auf den ersten zwei bis drei Seiten einer Druckausgabe entspricht (Tab. 1). Das EAK enthält nur Texte, die vom *Deutschen Textarchiv* (DTA) als fiktionale Erzählliteratur eingestuft wurden. Dazu

⁸¹ Dabei soll die u. a. von Heinrich Detering und Kai Sina („Der deutschsprachige Roman 1900–1950“, in: Volker Meid (Hg.), *Geschichte des deutschsprachigen Romans*, Stuttgart 2013, 445–623) als konstitutives Element moderner Erzählliteratur beschriebene „strukturelle Vielfalt des Literatursystems“ abgebildet werden, von kritisch-reflektiven (und auch experimentellen) Texten bis zu realitätsaffirmierenden Unterhaltungs- und Trivialromanen.

wurden aus den verfügbaren Einträgen des *DTA*-Teilkorpus ‚Belletristik‘ (Version vom 8.7.2015)⁸² alle Einträge vor 1880 sowie einige weitere Texte entfernt.⁸³

Das *DTA* ist als ‚historisches Referenzkorpus‘ des deutschen Standards einem nicht unerheblichen Nutzerkreis bekannt und leicht zugänglich⁸⁴ und so auch außerhalb der *Digital Humanities* hoch anschlussfähig. Zudem liegen klare Kriterien der Textauswahl zugrunde: Es wurden i. d. R. Erstausgaben von solchen Texten verwendet, für die eine breite zeitgenössische Leserschaft angenommen werden kann,⁸⁵ wobei „u. a. die Bibliographien ausgewählter Literaturgeschichten, die Textauswahl des Deutschen Wörterbuchs („Grimmsches Wörterbuch“) sowie Empfehlungen der Mitglieder der BBAW als Spezialisten der verschiedenen Disziplinen“ zugrunde gelegt wurden.⁸⁶ Auch wenn eine etwas detailliertere Dokumentation dieser Kriterien angeregt sein soll, ist mit der gegebenen Stichprobe eine plausible Datengrundlage gewählt, die zudem frei zugänglich ist und so die Transparenz und Wiederholbarkeit der Analysen garantiert. Ein weiteres Kriterium für die Wahl des *DTA* war die Existenz einer auf Tokenisierung und orthographischer Normalisierung aufbauenden *Part-of-Speech*-Annotation,⁸⁷ die eine weitere externe Ressource für die Metaphernidentifikation darstellt.⁸⁸ Zusätzlich annotierten wir im Rahmen des Reliabilitätstests den Anfang von Ferdinand

⁸²Die Textauswahl erfolgte über die online durchsuchbare Liste unter <http://www.deutschestextarchiv.de/list>. Das *DTA* ist in verschiedenen Versionen zum Download verfügbar unter: <http://www.deutschestextarchiv.de/download> (letzter Aufruf der Links 8.5.2017).

⁸³Vgl. J. Berenike Herrmann, „Anschaulichkeit messen. Eine quantitative Metaphernanalyse deutschsprachiger Erzählanfänge zwischen 1880 und 1926“, in: Tilmann Köppe/Rüdiger Singer (Hg.), *Show, don't tell. Konzepte und Strategien narrativer Anschaulichkeit*, Bielefeld 2018, 167–212.

⁸⁴<http://www.deutschestextarchiv.de/doku/leitlinien> (letzter Aufruf 1.9.2017).

⁸⁵Da die Auswahl der Texte in erster Linie unter sprachwissenschaftlich-lexikographischen Gesichtspunkten erfolgte, sind vornehmlich solche Texte aufgenommen worden, die „den historischen Sprachstand möglichst unverfälscht“ darstellen – das *DTA* nutzt i. d. R. die erste gedruckte, selbstständige Publikation eines Textes.

⁸⁶<http://www.deutschestextarchiv.de/doku/textauswahl> (letzter Aufruf 1.9.2017).

⁸⁷Bryan Jurish, *Finite-state Canonicalization Techniques for Historical German*, PhD thesis, Potsdam 2012, <http://opus.kobv.de/ubp/volltexte/2012/5578/> (letzter Aufruf 1.9.2017).

⁸⁸Eine Genauigkeit der Wortartenkennung für den Untersuchungszeitraum wurde im Mittel auf knapp unter 90 % geschätzt, was unter dem gemeinhin für das Deutsche angenommenen Wert von 97 % liegt (vgl. J. Berenike Herrmann, „Praktische Tagger-Kritik. Zur Evaluation des POS-Tagging des Deutschen Textarchivs“, in: *Tagungsband Fünfte Internationale Tagung Digital Humanities im deutschsprachigen Raum [DhD]*, Köln 2018, 287–290). Zu berücksichtigen ist jedoch, dass die Messung auf einer Stichprobe beruht, die seltene Wortarten genauso gewichtet wie häufige. Zudem weicht die *DTA*-Belletristik als Nichtstandardvarietät historisch und qua Textsorte von den Sprachmodellen der so evaluierten Tagger ab, die auf Zeitungstexten der Gegenwart trainiert wurden.

von Saars *Schloss Kostenitz*, einer realistischen Novelle mit modernen Zügen von 1893 (n = 236 Wörter).⁸⁹

4.2 Prozedur

Die Identifikation metaphorischen Sprachgebrauchs besorgten die Autorin (als Mitentwicklerin der ursprünglichen MIPVU für das Englische) und eine studentische Hilfskraft (die MIPVU u. a. durch ihre Bachelorarbeit eingeübt hatte). Das als CSV-Datei aufbereitete EAK wurde zwischen Oktober 2016 und März 2017 auf MRWs annotiert, wobei die *Part-of-Speech-Codes* als Informationsquelle insbesondere bei der Entscheidung über abgetrennte Partikeln komplexer Verben berücksichtigt wurden. Als externe Ressource wurde der *Duden* als korpusbasiertes Wörterbuch systematisch für jedes Wort konsultiert, eine ‚Zweitmeinung‘ lieferte das DWDS. Die beiden Annotatorinnen trafen sich in regelmäßigen Sitzungen, teilweise auch mit einer zweiten studentischen Hilfskraft, um Zweifelsfälle und Entscheidungen zu diskutieren. Alle EAK-Texte wurden von der Autorin (co-)annotiert.⁹⁰

4.3 Ergebnisse

Die statistische Auswertung zeigt eine signifikante Assoziation zwischen den Faktoren ‚Text‘ und ‚Metapher‘,⁹¹ der Mittelwert metaphorisch gebrauchter Wörter liegt bei 14,1 % (Standardabweichung = 4,9) (siehe Tab. 2).⁹² Abb. 1

⁸⁹Zugrunde liegt die Ausgabe im Georg Weiß Verlag, Heidelberg 1893, <https://www.projekt-gutenberg.org/saar/kosteniz/kosteniz.html> (letzter Aufruf 1.7.2022).

⁹⁰Ein Teil der Texte wurde von der Autorin vor-, von der SHK gegen- und von der Autorin abschlussannotiert (n = 11), ein zweiter Teil durch beide unabhängig annotiert und die Ergebnisse abgeglichen (n = 3), ein dritter Teil wurde durch die Autorin allein annotiert (n = 21).

⁹¹Um eine statistische Assoziation zwischen Metapherngebrauch und dem Faktor ‚Text‘ zu testen, wurde ein χ^2 -Test (sprich chi-Quadrat) durchgeführt. Verwendet wurde R (Version 3.3.3) mit den im Package lsr enthaltenen Funktionen `chisq.test(x)` und `cramersV(x)`: Daniel J. Navarro, „Learning Statistics with R: A Tutorial for Psychology Students and other Beginners“ (Version 0.5), Adelaide, Australia, 2015, <https://cran.r-project.org/web/packages/lsr/index.html> (letzter Aufruf 1.9.2017). Das Ergebnis ist $\chi^2 = 363,21(34)$, $p < 0,001$, Cramer's V = 0,14. Es sei darauf hingewiesen, dass die Effektstärke, die hier durch Cramer's V gemessen wird, mit 0,14 im Bereich einer sogenannten „small effect size“ liegt. Vgl. Jacob Cohen, *Statistical Power Analysis for the Behavioral Sciences*, Hillsdale, NJ ²1988, 25 und 79, vgl. auch Stefan Th. Gries, „Some Current Quantitative Problems in Corpus Linguistics and a Sketch of Some Solutions“, in: *Language and Linguistics* 16/1 (2015), 93–117.

⁹²Der χ^2 -Test ist in der Korpuslinguistik umstritten, da er die Unabhängigkeit der gemessenen Fälle zugrunde legt, Sprache aber *per definitionem* aus syntagmatischen Folgen besteht. Vgl. z. B. Stefan Evert, „How Random is a Corpus? The Library Metaphor“, in: *Zeitschrift für Anglistik und Amerikanistik* 54/2 (2006), 177–190; Adam Kilgariff, „Language is Never, Ever, Ever, Random“, in: *Corpus Linguistics and Linguistic Theory* 1/2 (2005), 263–276. Für die vorliegende Studie mit einer vergleichsweise geringen Fallzahl bietet sich jedoch ein Modus an,

Tab. 2 Relative Häufigkeiten MRWs/Non-MRWs pro Textanfang

TextID	AutorIn	Titel	Non-MRW	MRW	Gesamt
1	Altenberg, Peter	Prödrömös	73,25 %	26,75 %	100,00 %
2	Andreas-Salomé, Lou	Fenitschka. Eine Ausschweifung	86,18 %	13,82 %	100,00 %
3	Bierbaum, Otto Julius	Stilpe	85,23 %	14,77 %	100,00 %
4	Déry, Juliane	Selige Liebe	79,88 %	20,12 %	100,00 %
5	Eschstruth, Nataly von	Katz' und Maus	83,10 %	16,90 %	100,00 %
6	Fontane, Theodor	Schach von Wuthenow	83,00 %	17,00 %	100,00 %
7	Fontane, Theodor	Effi Briest	87,55 %	12,45 %	100,00 %
8	Fontane, Theodor	Der Stechlin	84,22 %	15,78 %	100,00 %
9	Fontane, Theodor	Irrungen, Wirrungen	84,45 %	15,55 %	100,00 %
10	Hagenauer, Arnold	Muspilli	77,98 %	22,02 %	100,00 %
11	Heyking, Elisabeth von	Die Orgelpfeifen	91,16 %	8,84 %	100,00 %
12	Kafka, Franz	Das Urteil	90,59 %	9,41 %	100,00 %
13	Kafka, Franz	Die Verwandlung	91,68 %	8,32 %	100,00 %
14	Kafka, Franz	Ein Hungerkünstler	87,23 %	12,77 %	100,00 %
15	Kafka, Franz	Der Prozess	90,33 %	9,67 %	100,00 %
16	Keller, Gottfried	Das Sinngedicht	80,78 %	19,22 %	100,00 %
17	Keyserling, Eduard von	Beate und Mareile	84,36 %	15,64 %	100,00 %
18	Kraft, Robert	Der Medizinmann	87,59 %	12,41 %	100,00 %
19	Kretzer, Max	Meister Timpe	76,00 %	24,00 %	100,00 %
20	Laßwitz, Kurd	Auf zwei Planeten	81,33 %	18,67 %	100,00 %
21	May, Karl	Durchs Wilde Kurdistan	91,26 %	8,74 %	100,00 %
22	Meißner, Alfred	Die Prinzessin von Portugal	84,45 %	15,55 %	100,00 %
23	Polenz, Wilhelm von	Der Büttnerbauer	85,33 %	14,67 %	100,00 %
24	Preuschen, Hermione von	Yoshiwara	81,74 %	18,26 %	100,00 %
25	Raabe, Wilhelm	Stopfkuchen	82,31 %	17,69 %	100,00 %
26	Raabe, Wilhelm	Die Akten des Vogelsangs	82,92 %	17,08 %	100,00 %
27	Reventlow, Fanny Gräfin zu	Herrn Dames Aufzeichnungen	82,39 %	17,61 %	100,00 %
28	Schnitzler, Arthur	Fräulein Else	94,21 %	5,79 %	100,00 %
29	Schnitzler, Arthur	Traumnovelle	92,84 %	7,16 %	100,00 %
30	Schreiner, Olive	Peter Halket im Mashonalande	91,91 %	8,09 %	100,00 %
31	Schublin, Ossip	Vollmondzauber	90,15 %	9,85 %	100,00 %
32	Spyri, Johanna	Heidi's Lehr- und Wanderjahre	90,25 %	9,75 %	100,00 %
33	Storm, Theodor	John Riew'	87,50 %	12,50 %	100,00 %
34	Storm, Theodor	Ein Doppelgänger	89,42 %	10,58 %	100,00 %
35	Sturza, Marie Tihanyi	Das Gelübde einer dreißigjährigen Frau	92,41 %	7,59 %	100,00 %

demonstriert, dass ungefähr die Hälfte der Texte ($n=17$ von $N=35$) eine durchschnittliche Metaphernfrequenz besitzt, während der Rest entweder positiv oder negativ vom statistisch erwarteten Wert abweicht. Die Abbildung zeigt für jeden Textanfang (x -Achse) die standardisierten Residuen für MRWs (y -Achse). Diese geben Hinweise auf den Grad der Abweichung der gemessenen Metaphernhäufigkeit von einem statistisch geschätzten (Erwartungs-)Wert.⁹³ Vereinfacht

der nicht im engeren Sinne hypothesenüberprüfend vorgeht, sondern das statistische Modell zur Musterexploration, vgl. auch Rainer Perkuhn/Holger Keibel/Marc Kupietz, *Korpuslinguistik*, Paderborn 2012.

⁹³ Standardisierte Residuen bilden diese Abweichung in Einheiten der Standardabweichung ab. Sie sind zunächst ein Mittel, um zu bestimmen, welche Zellen der Kreuztabelle (hier: MRWs/NonMRWs und Textfragment) zum signifikanten χ^2 -beitragen. So kann man sie gebrauchen, um Muster in den Ergebnissen aufzufinden: Werte über/unter $\pm 2,58$ zeigen eine signifikante Abweichung vom erwarteten Wert. Das gewählte Signifikanzniveau setze ich aufgrund der kleinen Effektgröße bei 1 % Wahrscheinlichkeit des α -Fehlers an (in Abb. 1 durch die gestrichelte Linie dargestellt; Werte über/unter $\pm 1,96$ haben eine α -Fehlerwahrscheinlichkeit von 5 %, durchgezogene Linie).

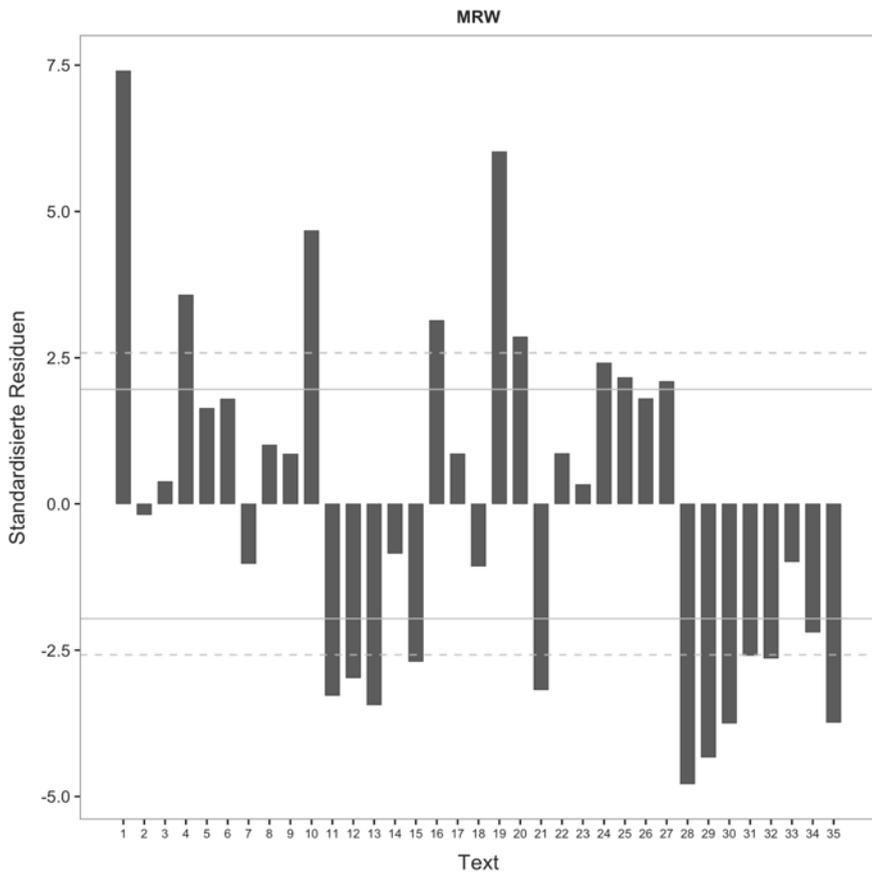


Abb. 1 EAK. Standardisierte Residuen MRWs pro Textanfang

gesagt, haben diejenigen Texte im Bereich zwischen den gestrichelten Linien (zwischen ± 2.58) durchschnittliche Werte; diejenigen Texte, die darüber hinausragen, sind hingegen signifikant metaphernreicher oder -ärmer (vgl. Tab. 2, die die relativen Häufigkeiten der metaphorisch gebrauchten Wörter für die einzelnen 35 Textanfänge ausweist, wobei der Anteil der MRWs gegenüber den non-MRWs pro Text dargestellt wird: Die minimale Proportion zeigt mit 6 % MRWs Schnitzlers *Fräulein Else*, die maximale mit ca. 27 % MRWs Altenbergs *Prödrömös*).

Ein Vergleich mit dem Englischen *VUAMC* kann die beobachteten Häufigkeiten etwas einordnen, lässt doch die Standardisierung der Methode (unter der gebotenen Vorsicht bei Vergleichen über Sprachen und Korpora hinweg!) diesen Vergleich zumindest heuristisch-informell zu. Im *Fiction-Sample* sind 11,9 % der

Wörter MRWs, im *News-Sample* 16,4 %, während die akademischen Texte mit 18,5 % besonders metaphernreich und Gespräche mit 7,7 % besonders metaphernarm sind.⁹⁴ Mit 14,1 % liegt das EAK also etwas über der englischen *Fiction*.

Hier sind zunächst der Englisch-Deutsch-Sprachkontrast und die historische und Gattungsdifferenz der Korpora zu berücksichtigen, enthält das *Fiction-VUAMC* im Unterschied zum Jahrhundertwende-EAK doch durchgängig populäre Romane der 1990er Jahre, die Dorst mit Jakobson als vergleichsweise metonymisch einordnet. Doch auch die Tatsache, dass das EAK nur aus *Erzählanfängen* besteht, könnte mitverantwortlich für die Unterschiede sein – und die proportionale Ähnlichkeit zu den englischen Zeitungstexten eine Strategie der Aufmerksamkeitssteigerung durch vermehrte metaphorische Elemente im Exordium anzeigen.

Die Verteilung der MRWs über die Hauptwortarten lautet wie folgt: Die relative Häufigkeit von MRW-Präpositionen ist mit 48 % im EAK deutlich höher als im *Fiction-Sample* des *VUAMC* (33,4 %), ja übersteigt sogar den durchschnittlichen Wert für das *VUAMC*-Gesamtkorpus (38 %). Auch der Wert für metaphorische Adjektive übersteigt mit 23 % den Wert des *Fiction-Samples* (19,4 %), ebenso wie der für metaphorische Verben mit 21,5 % (15,9 %). Der Wert für Nomen entspricht mit 11,1 % dagegen in etwa dem des *Fiction-Samples* (10,5 %), und nur der Anteil der Adverbien ist mit 6,4 % niedriger als der des englischen Roman-korpus (9,3 %). Eine ausführlichere Analyse der metaphorischen Wortartenverteilung kann an dieser Stelle zwar nicht geleistet werden,⁹⁵ doch mag spekuliert werden, dass das Deutsche mit seinen längeren Sätzen beispielsweise generell mehr metaphorische Präpositionen zur Einbettung von Subordinationen nutzt; es ist aber denkbar, dass hier der historische Abstand oder die Zusammensetzung des EAK zu Buche schlägt.

Interessanter ist, dass das EAK drei Hauptcluster bildet: durchschnittliche, metaphernreiche, und metaphernarme Texte.⁹⁶ Die siebzehn ‚durchschnittlichen‘ Texte sind stilistisch und auch inhaltlich heterogen, es sind moderne Schreibweisen neben realistischen zu finden, aber auch ‚ernste‘ Literatur neben sentimental-trivialer und unterhaltender. Sechs Texte sind besonders metaphernreich (ca. 19–27 %): Altenbergs *Pròdròmös*, Max Kretzers *Meister Timpe*, Hagenauers *Muspilli* und Dérys *Selige Liebe* sind dabei modernistische Texte,

⁹⁴ Steen/Dorst/Herrmann u. a. (Anm. 7).

⁹⁵ Vgl. Herrmann (Anm. 82).

⁹⁶ Ebd.

in denen MRWs u. a. zur Schilderung subjektiver Weltsicht in Bezug auf innere Affekte und Impressionen des Äußeren genutzt werden, während Kellers *Das Sinnedicht* und Laßwitz' Science Fiction *Auf zwei Planeten* metaphorischen Sprachgebrauch im Rahmen realistischer Erzählmittel ausschöpfen.

Besonders wenig MRWs zeigen zum einen Erzählanfänge der Unterhaltungs- und Kolportageliteratur von weiblichen Autoren: Heykings *Orgelpfeifen* (1918), Spyris *Heidi* (1880), Schubins *Vollmondzauber* (1899), Sturzas *Gelübde* (1905) und Schreiners *Peter Halket* (1898). May mit *Kurdistan* (1892) fällt als einziger männlicher Autor auf (vgl. Sprengels Befund der Nähe Mays zum „sentimental-trivialen Frauenroman“).⁹⁷ Diese Texte berichten in ‚realistischem‘ Duktus von konkret-physischen und sprachlichen Handlungen in situierter Umgebung. Ihre Schilderung konkreter Dinge, Körper und Vorgänge kommt wohl mit weniger metaphorischen Ausdrücken aus; mitverantwortlich kann aber auch eine geringere Satzkomplexität sein, die nur wenige konventionalisierte Metaphern für ‚textuelle‘ Funktionen braucht. Die zweite Untergruppe der metaphernarmen Texte versammelt Schnitzlers *Fräulein Else* (1924) und *Traumnovelle* (1926) sowie Kafkas *Das Urteil* (1913) und *Der Prozess* (1925), die literaturhistorisch wohl den „Pionier- und Meisterwerken“ einer „reflektierten Moderne“ zugeordnet werden⁹⁸ und deren verbindendes Element hier ist, dass auf der narrativen Oberfläche Handlungen, Beschreibungen der externen/gewünschten Welt sowie Dialoge vorherrschen.

Die quantitative Untersuchung des EAK zeigt Unterschiede und Gemeinsamkeiten des Metapherngebrauchs, für deren Erklärung die Faktoren Autorstil⁹⁹ und Gattung sowie nicht zuletzt literaturepochale Aspekte fruchtbare Ansatzpunkte sind. So weist der *Korrelationskoeffizient* eine signifikante Korrelation von MRWs mit dem Faktor ‚Zeit‘ (Publikationsdatum) aus ($r = -0,34$ (35), $p < 0,01$).¹⁰⁰ Abb. 2 zeigt die Korrelationsgerade und das Streudiagramm für den Zusammenhang.

Die signifikante Korrelation könnte, wenn man den überschaubaren Datensatz vorsichtig interpretiert, andeuten, dass sich über die sechsendreißig Jahre erstreckende Stichprobe hinweg eine reflektierte Moderne hin zu Kafka und Schnitzler bei Abnahme von figurativem Sprachgebrauch entfaltet. Eine

⁹⁷ Peter Sprengel, *Geschichte der deutschsprachigen Literatur 1900–1918: Von der Jahrhundertwende bis zum Ende des Ersten Weltkriegs*, München 2004, 203.

⁹⁸ Helmuth Kiesel, *Geschichte der deutschsprachigen Literatur 1918 bis 1933*, München 2017, 1173 ff.

⁹⁹ Texte von Autoren, die mit mehr als einem Text vertreten sind, sind jeweils klar einer der drei Gruppen zuzuordnen: Kafkas und Schnitzlers Erzählanfänge gehören (bis auf *Ein Hungerkünstler*, der das Signifikanzniveau nicht erreicht) zu den metaphernarmen Texten, während Fontanes Texte, ebenso wie Raabes und Storms, qua Metapherndichte zu den ‚durchschnittlichen‘ zählen.

¹⁰⁰ Zur Errechnung der Korrelationen sowie ANOVAs wurde R (version 3.3.3) mit den Funktionen *cor(x)* und *aov(x)* verwendet, sowie die Funktion *biserial.cor(x)* der *ltm*-Package (Dimitris Rizopoulos, „ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses“, in: *Journal of Statistical Software*, 17/5 [2006], 1–25, <http://www.jstatsoft.org/v17/i05/> [letzter Aufruf 1.9.2017]).

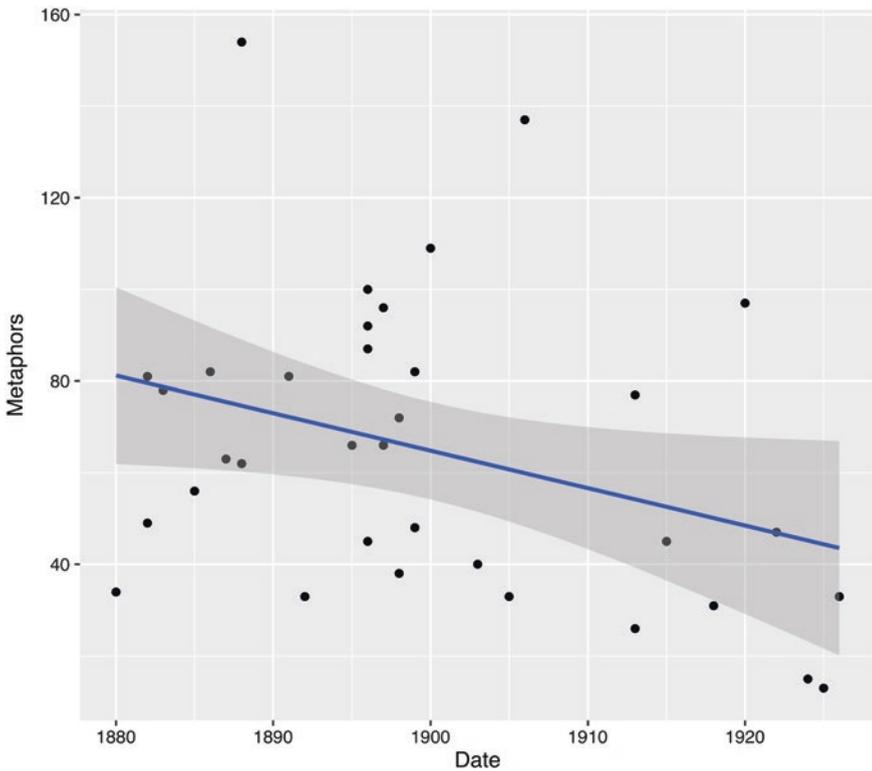


Abb. 2 Korrelation Metapher (MRW) und Zeit (Publikationsdatum)

empirische Frage ist jedoch, ob sich hier nicht ein allgemeiner sprachgeschichtlicher Trend abbilden könnte. Ungeklärt ist auch, wie sich ‚kreative‘ zu ‚konventionellen‘ Metaphern verhalten. Auch ob und wie hier Faktoren wie Autorstil, Handlung o. ä. mitbestimmen und ob es sich nicht eher um ein umgedrehtes ‚U‘ (einen nicht-linearen Zusammenhang) handelt – von Jakobsons eher metonymischem Realismus über den metaphorischen Symbolismus zur nicht-metaphorischen reflektierten Moderne –, muss in weiteren Studien (mit einer größeren Stichprobe) bestimmt werden.

Weitere statistische Interaktionen zwischen ‚Metapher‘ und anderen untersuchten Variablen waren nicht signifikant. Es fielen aber andere signifikante Inferenzstatistiken auf, die kurz zusammengefasst werden sollen: Ein *Chi-Quadrat*-Test zeigte eine Assoziation zwischen *Flesch-Index* und dem Faktor ‚Text‘ ($\chi^2=105,53(34)$; $p<.001$), die einzelnen Texte unterscheiden sich also signifikant bezüglich der so gemessenen morpho-syntaktischen Komplexität. Eine Richtung könnte in der diachronen Dimension liegen, so korrelieren Satz- ($r=-.19$) und Wortlänge ($r=-.23$) jeweils negativ mit ‚Zeit‘ (sie nehmen also

ab), während der *Flesch-Index* (als Maß morpho-syntaktischer Komplexität)¹⁰¹ ansteigt ($r = .20$).¹⁰²

Zudem zeigt eine Analyse der Wortartenverteilung, dass die Erzählanfänge über die Zeit hinweg weniger Nomen ($r = -.34$), Präpositionen ($r = -.44$) und Artikel ($r = -.33$) aufweisen, dafür aber mehr Verben ($r = .26$), Adverbien ($r = .37$) und Pronomen ($r = .35$). Dies weist auf eine mögliche Entwicklung zu einem kolloquialen, informelleren Stil hin,¹⁰³ die aber auch außerhalb von literarischen Text(anfäng)en plausibel ist.

Schließlich zeigt eine Varianzanalyse (ANOVA) einen Interaktionseffekt von ‚narrativer Perspektive‘, *Flesch-Index* und ‚Zeit‘ ($p = .015$), was bedeutet, dass die Erzählanfänge in vornehmlich erster Person eher kurze Sätze und kürzere Wörter verwenden als jene in vornehmlich dritter Person (siehe Abb. 3). Der informellere Ton könnte in der ersten Erzählperson stilprägender sein. All diese Beobachtungen sind vielversprechend und sollen an einem Korpus, das die Erzählanfänge mit den Haupt- und Schlussteilen der Texte vergleicht, weiter überprüft werden.

Aus der Perspektive der digitalen Literaturwissenschaft stehen die spezifischen Merkmale des *literarischen* Metapherngebrauchs im Vordergrund. Der gewählte Ansatz zielt, gerade durch die Modellierung der Metapher als graduelles und kontextsensitives Phänomen zwischen kreativem und konventionellem Gebrauch, explizit auf den Vergleich mit nichtliterarischen Erzähltexten (bzw. deren Anfängen) desselben Zeitraums ab. Metaphern sind, wie am EAK gezeigt werden konnte, tatsächlich durch höhere Grade von Ambiguität und Ästhetisierung gekennzeichnet und werden für unterschiedliche erzählerische Zwecke eingesetzt.¹⁰⁴ Aus methodischer Sicht ist ambige Metaphorizität ein Grenzfall, der entweder durch MIPVU selbst oder durch *Coder-Disagreement* im Reliabilitätstest identifiziert werden kann. Diese Fälle können quantitativ analysiert und heuristisch genutzt werden, um stilistisch interessante Fälle zu lokalisieren.

MIPVU enthält einen Code zur Ausflagung von Grenz- oder Zweifelsfällen, das sogenannte ‚WIDLII‘ (‚When In Doubt Leave It In‘).¹⁰⁵

Im EAK, das insgesamt $N = 18.092$ Wörter umfasst, erhielten $n = 790$ aller Wörter den Code ‚WIDLII‘, was 4,3 % der Gesamtwortzahl entspricht (siehe Tab. 3) und proportional höher als im *VUAMC* (0,9 %) ist. Unter den MRWs (Gesamtanzahl $n = 2469$) entspricht dies dann immerhin einem Anteil von 29,9 % (siehe Abb. 4).

¹⁰¹ Vgl. Bailin/Grafstein (Anm. 16).

¹⁰² Der *Flesch-Index* wurde in R (Version 3.3.3) mit der Funktion *readability(x)* der *koRpus-Package* errechnet (Meik Michalke, „koRpus. An R Package for Text Analysis“ (Version 0.10–2, 2017), <https://reaktanz.de/?c=hacking&s=koRpus> (letzter Aufruf 1.9.2017)).

¹⁰³ Vgl. Biber (1988) et passim (Anm. 18).

¹⁰⁴ Herrmann (Anm. 82).

¹⁰⁵ Steen/Dorst/Herrmann u. a. (Anm. 48), 33 f.

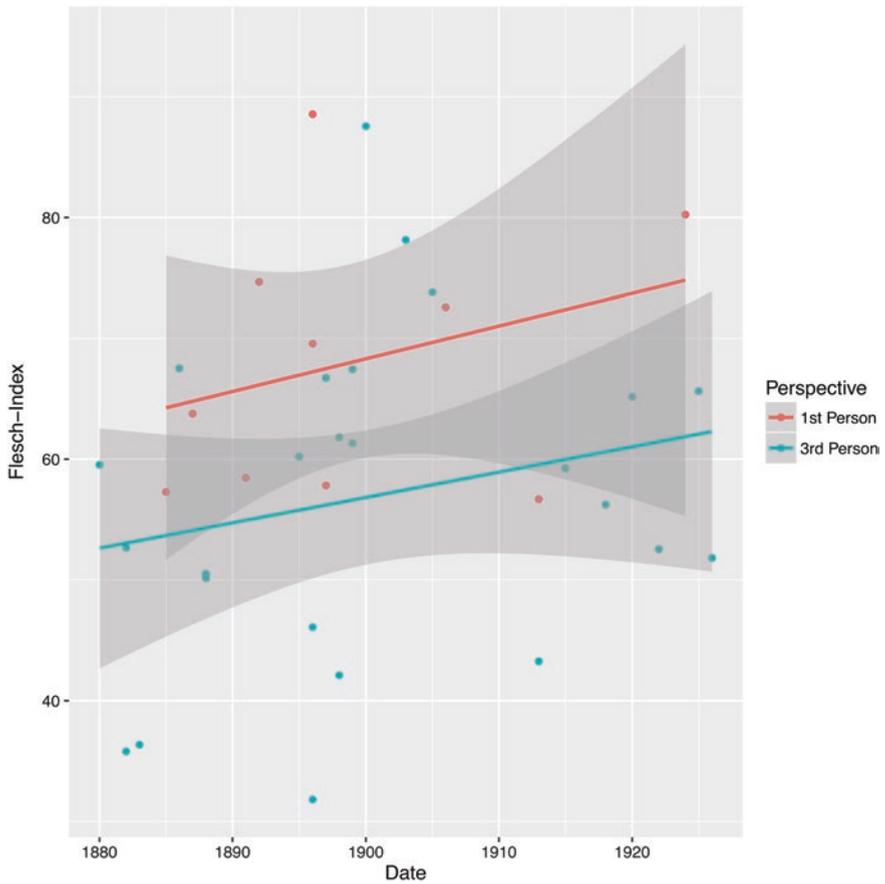


Abb. 3 Interaktion *Flesch-Index*, Zeit (Publikationsdatum) und Erzählperspektive

Tab. 3 Absolute und relative Häufigkeiten, Mittelwert und Standardabweichung pro MRW-Typ

	Non_MRW	MRW_Clear	MRW_WIDLII	MRW_gesamt
Summe	15.623	1679	790	2469
% EAK	86,4 %	9,3 %	4,3 %	13,6 %
Mittelwert	446,37	47,97	22,57	70,54
Std. Abw	32,13	15,54	11,49	23,00

Die WIDLII-Fälle sind solche, bei denen entweder die Struktur des Sprachsystems oder die Struktur des Wörterbuches eine klare Identifizierung erschwert, sie sollen also weiter untersucht und zur Validierung der MIPVU für das Deutsche genutzt werden. Dazu kommen aber auch Fälle, in denen aufgrund einer Bedeutungsambiguität im Kontext nicht klar entschieden werden konnte,

EAK: MRW-Typen

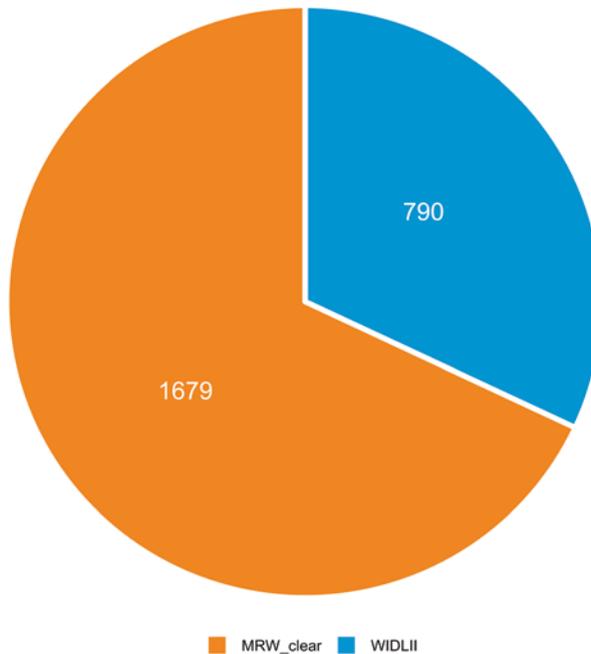


Abb. 4 Anteil der WIDLIIs an allen metaphorisch verwendeten Wörtern im EAK in Prozent

ob es sich um einen Kontrast bei gleichzeitiger Ähnlichkeit der Bedeutungen eines Wortes handelt. Solche Fälle sind oft stilistisch besonders interessant. So trat zum Beispiel für den Ausdruck „sonniges“ (4) im Reliabilitätstest ein *Coder-Disagreement* auf, was anzeigt, dass der Status unklar ist. In der Tat zeigt eine nähere Analyse, dass hier *gleichzeitig* eine metaphorische („von einer offenen, freundlichen Wesensart; heiter“) und eine buchstäbliche („von der Sonne beschienen“) Bedeutung vorhanden sind:

- 4 Unten aber dehnte sich die Ebene aus, damals wie heute, ein *sonniges* Bild regen, werkhätigen Lebens. (Ferdinand von Saar, *Schloss Kostenitz*, 1893)

Von Saar, ein österreichischer Autor des Realismus, gilt als „Figur des Übergangs zum Fin du Siècle von Hoffmannsthal, Schnitzler und Bahr“, der die „Brüche und Dissonanzen“ der franzisko-josephinischen Epoche „mitunter durch konventionelle Erzählschlüsse entschärft“, zum Beispiel durch Idyllen.¹⁰⁶ Eine scheinbar konventionelle Idylle findet sich auch im Erzählanfang von *Schloss Kostenitz*, wo gerade das MRW „sonnig“ in (4) eine konventionell-realistische Figurativität

¹⁰⁶Vgl. Karl Wagner, „Saar, Ferdinand Ludwig Adam von“, in: *Neue Deutsche Biographie (NDB)* 22 (2005), 315 f., hier: 316.

evoziert, die im Sinne Jakobsons zwischen Metapher und Metonymie steht.¹⁰⁷ Ähnlich liegt in Raabes *Die Akten des Vogelsangs* (5) eine zugleich metaphorische und wörtliche Verwendung des Nomen „Leere“ vor, das hier sowohl auf ‚physische‘ als auch ‚emotionale Einsamkeit‘ referiert. Auch hier waren sich die Annotatorinnen uneinig. Die Ich-Erzählerin berichtet vom Tod des Protagonisten Velten in einem Brief, den sie in seiner Wohnung schreibt:

- 5 [Velten] ist allein geblieben bis zuletzt, mit sich selber allein. [...] Es *klingt hohl in dem Raume, in welchem ich schreibe*: Er hat die *Leere um sich gelassen* [...]. (Wilhelm Raabe, *Die Akten des Vogelsangs*, 1896)

Der Diskurs in (5) zeigt eine Rekurrenz von Wörtern, die sich auf Räumliches in einer Doppelbedeutung beziehen – so befindet sich die Briefschreiberin in einem leeren Raum, der zugleich metaphorisch und metonymisch für die Distanziertheit und Einsamkeit seines ehemaligen Bewohners steht. Der ambige Status zwischen wörtlicher und metaphorischer Bedeutung ist so ein potenzieller Anlass für erhöhte leserseitige Aufmerksamkeit, die sich gegebenenfalls auch auf die Präpositionen („in“, „um“) bezieht. Literarizität als Kontextfaktor trägt, so ist anzunehmen, zur hintergründigen Anschaulichkeit der MRW „Zurückgezogenheit“ in (6) bei, das ebenfalls ein *Coder-Disagreement* zu verzeichnen hatte:

- 6 [...] eine große, *feldeinwärts* sich erstreckende Gärtnerei, deren kleines, dreifenstriges, in einem Vorgärtchen *um* etwa hundert Schritte zurückgelegenes Wohnhaus, trotz aller Kleinheit und *Zurückgezogenheit*, von der vorübergehenden Straße her sehr wohl erkannt werden konnte. (Theodor Fontane, *Irrungen, Wirrungen*, 1888)

Das personifizierende „Zurückgezogenheit“ ist typisch für Fontanes Stil, durch häufige Verwendung subtiler Animierung und Personifizierung die Beschreibung von Unbelebtem zu dynamisieren. Das Nomen ist lexikalisiert mit einer sich auf menschliches Verhalten beziehenden Grundbedeutung ‚Zustand des Sich-zurückgezogen-Habens, Abgeschiedenheit, Kontaktlosigkeit‘.¹⁰⁸ Diesen Eindruck bestätigt nicht nur das korpusbasierte Beispiel im *Duden* („in [völliger] Zurückgezogenheit leben“), sondern auch das konsultierte *DWDS*, welches das Nomen als Ableitung des partizipialen Adjektivs *zurückgezogen* definiert.¹⁰⁹ Dessen Grundbedeutung ‚den Kontakt mit den Menschen seiner Umgebung, gesellschaftlichen Umgang meiden, ganz für sich leben‘ kann klar ein menschliches Agens zugewiesen werden (die Anwendung auf das Haus ist personifizierend-metonymisch). Ob die Verwendung von *Zurückgezogenheit* tatsächlich zu gesteigerter Dynamisierung im Leserleben führt, ist allerdings eine empirische Frage.¹¹⁰

¹⁰⁷ Jakobson (Anm. 15), 135.

¹⁰⁸ <http://www.duden.de/rechtschreibung/Zurueckgezogenheit> (letzter Aufruf 1.9.2017).

¹⁰⁹ <https://www.dwds.de/wb/Zurueckgezogenheit> (letzter Aufruf 1.9.2017).

¹¹⁰ Die leserseitigen Faktoren müssen empirisch auf der Ebene des *Reader Response* geklärt werden. Siehe etwa Hortons Studie zu metaphorischen Ausdrücken in der Wahrnehmung sozialer Nähe in kurzen Narrativen: William S. Horton, „Metaphor and Readers’ Attributions of Intimacy“, in: *Memory & Cognition* 35/1 (2007), 87–94.

5 Fazit

Ausgehend von der Beschreibbarkeit von Regeln der literarischen Bedeutungskonstitution sowie von der besonders augenfälligen Eignung der Metapher zur literarischen Verfremdung wurde im vorliegenden Aufsatz die gängige literaturwissenschaftliche Perspektive auf Metaphern gegen den Strich gebürstet. Als Untersuchungsgegenstand modelliert wurden nicht nur verfremdende, sondern die gesamte potenzielle Bandbreite metaphorischer Ausdrücke. Dabei wurde MIPVU zur Operationalisierung als eine datenbasierte Methode des manuellen *Distant Reading* mit dem Ziel der Beobachterunabhängigkeit, Reliabilität und Vergleichbarkeit von Analysen eingeführt. Im Unterschied zur hermeneutischen Metaphorologie wird hier die Interpretation in eine Forschungsphase aufgeschoben, die erst nach kontrollierter Identifikation und statistischer Analyse erfolgt.

Ambivalente und strittige Fälle werden dabei in der Phase der Annotation markiert und können so in weiteren quantitativen wie qualitativen Schritten näher untersucht und kategorisiert werden – wie etwa „sonnig“ und „Leere“, die im jeweiligen literarischen Kontext als doppelt codiert, also zugleich metaphorisch und eigentlich, auffielen.

Der Hauptnutzen der vorgestellten Operationalisierung ist eine Externalisierung und Standardisierung des Vorgehens, die einen Vergleich über Textpassagen, Texte, Gattungen und Register oder mit gewissen Abstrichen auch über Sprachen hinweg ermöglicht.¹¹¹ Gerade deshalb weist MIPVU auch Metaphern aus, die aus literaturwissenschaftlicher Perspektive oftmals übersehen werden. Durch die formale Modellierung können Metaphern, wie gezeigt, auf ihre statistisch messbare Interaktion mit Faktoren wie Wortarten und morphosyntaktischer Komplexität (*Flesch-Index*), aber auch narratologischen Operationalisierungen wie Erzählperspektive in diachroner Perspektive untersucht werden. Beobachtet werden konnte hier eine Tendenz zur Abnahme metaphorisch verwendeter Wörter im Übergang vom Realismus zur Moderne, ebenso wie eine Abnahme der morphosyntaktischen Komplexität – beide Trends stoßen weitere Fragen an.

Externe Validität im Sinne einer möglichst einzelbeobachterunabhängigen Annotation wird mit MIPVU durch die regelgeleitete Einbindung von korpusbasierten Wörterbüchern angestrebt. Dabei wird die kontextuelle Bedeutung eines Wortes bezüglich seiner Metaphorizität beurteilt: Kann sie mit einer ‚grundlegenderen‘ Bedeutung verglichen werden? Ist die Beziehung durch einen semantischen Kontrast geprägt? Die ‚Grundbedeutung‘ als die ‚eigentliche‘ Bedeutung des metaphorisch gebrauchten Wortes,¹¹² ist als erfahrungsnähere, körperorientierte, mit den Sinnen erfahrbare Bedeutung in der Regel im Wörterbuch zu finden. Hier kommt die

¹¹¹Susan Nacey/Aletta G. Dorst/Tina Krennmayr u. a. (Hg.), *MIPVU in Multiple Languages*, Amsterdam (2019) Zur Externalisierung des Lesens als geisteswissenschaftliche Methode siehe Herrmann (Anm. 2).

¹¹²Vgl. Rüdiger Zymner, „Uneigentliche Bedeutung“, in: Jannidis/Lauer/Martínez u. a. (Anm. 4), 128–168.

kognitionslinguistische Metapherdefinition zum Tragen, die aus der antiken Rhetorik¹¹³ wie aus den Theorien von Vico, Kant, Weinrich und Hermann Paul motiviert werden kann.¹¹⁴

Die interne Validität wird hingegen zum einen durch das iterative Annotationsprotokoll angestrebt, das Diskussionsrunden für jeden Durchgang vorsieht,¹¹⁵ zum anderen durch Reliabilitätstests, die die Verlässlichkeit, aber auch die Kalibrierung der Methode überprüfen. Ein Nachteil der gewählten Operationalisierung ist sicherlich die etwas geringere Trennschärfe für die Quantifizierung unterschiedlicher Typen literarischer Metaphern, die jedoch durch die Entwicklung neuer Kategorien ausgeglichen werden kann. So schlägt Steen die „deliberate metaphor“ vor, um semiotisch und kommunikationstheoretisch die Aktualisierung einer potenziellen Metaphorizität zu identifizieren, das Konstrukt ist jedoch nicht unumstritten.¹¹⁶ Andere Zugänge setzen bei einer merkmalsbasierten computationellen Analyse an, die mittels *Machine Learning* Leseratings miteinbezieht.¹¹⁷

Ein wichtiges Argument für die Erhebung der ‚gewöhnlichen‘ Metaphern ist, dass textuelle Metaphorizität ausschlaggebend durch gemeinhin häufig übersehene Textmerkmale wie Präpositionen und Adverbien, sowie verblichene und subtil-personifizierende Bedeutungen gestaltet wird. Untersucht man gerade diese Wörtchen, kann lokale und auch diskursive Kohärenz systematisch erfasst werden, etwa in bestimmten literarischen Kontexten, wo Phänomene wie *Fictive Motion*¹¹⁸ statische Begebenheiten dynamisieren („der Zaun führt nach Osten“). Wichtiger noch: nur wenn auch konventionelle Metaphorizität durch Annotation lernbar ist, können Verfahren des maschinellen Lernens in Anschlag gebracht werden, die nuancierte Metaphernidentifikation und -analyse über Gattungen, Epochen und Diskurse hinweg in *Big* oder *Small Literary Data* valide automatisieren.

Dass die Analyse der MIPVU-Daten agnostisch gegenüber der Dimension des empirischen Lesers¹¹⁹ bleibt und Metaphorizität als *semiotisches Potenzial* betrachtet, ist ein weiterer wichtiger Unterschied zu gängigen Spielarten der hermeneutischen Metaphorologie. Hier ist das Konstatieren der Potenzialität eine

¹¹³Vgl. Lausberg (Anm. 28).

¹¹⁴Vgl. auch Semino/Steen (Anm. 7).

¹¹⁵Steen/Dorst/Herrmann u. a. (Anm. 48).

¹¹⁶Gerard J. Steen, „Deliberate Metaphor Affords Conscious Metaphorical Cognition“, in: *Cognitive Semiotics* 5/1–2 (2009), 179–197. Vgl. z. B. Raymond W. Jr. Gibbs, „Are Deliberate Metaphors Really Deliberate? A Question of Human Consciousness and Action“, in: *Metaphor and the Social World* 1 (2011), 26–52.

¹¹⁷Bislang liegen Ergebnisse für Einzelausdrücke wie „doubt is a sword“ vor, vgl. Arthur M. Jacobs/Annette Kinder, „What Makes a Metaphor Literary? Answers from Two Computational Studies“, in: *Metaphor and Symbol* 33/2 (2018), 85–100.

¹¹⁸Teenie Matlock, „Metaphor, Simulation, and Fictive Motion“, in: Barbara Dancygier (Hg.), *The Cambridge Handbook of Cognitive Linguistics*, Cambridge, UK, 2017, 477–490.

¹¹⁹Steen schlägt vor, eine Debatte der Metaphernforschung zu entschärfen, indem man auch ‚wörtlich verarbeitete‘ konventionelle Ausdrücke als Metaphern beschreibt, solange sie als konzeptuelles *Mapping* semiotisch nachgewiesen werden können, vgl. Gerard J. Steen (Anm. 44).

Einladung zur arbeitsteiligen Forschung, etwa um Kategorien wie ‚Kreativität‘ oder ‚Anschaulichkeit‘ zu operationalisieren und eben auch das empirische Leserverhalten zu überprüfen: Wann und von wem werden welche Metaphern wie verstanden? Welche kognitiven und affektiven Rezeptionsprozesse werden angestoßen?

Die hier vorgestellte Operationalisierung der Metapher steht im Paradigma der algorithmischen Automatisierung der Metaphernidentifikation.¹²⁰ Für eine Ergänzung und Weiterentwicklung einer algorithmisch erweiterten Metaphorologie sind derzeit besonders vielversprechend Verfahren des maschinellen Lernens, die transfer-learning auf Transformer-Sprachmodellen wie BERT und GTP, Methoden der *Distributionellen Semantik*¹²¹ und *Neurale Netzwerke*¹²² anwenden. Solche Verfahren setzen freilich größere Datenmengen und externalisierende Perspektiven voraus, denen jedoch auch die literaturwissenschaftliche Metaphernforschung – nicht nur mit jeder validierbaren Tiefenannotation – stets näher rückt.¹²³

Literatur¹²⁴

- Bailin, Alan/Grafstein, Ann, *Readability. Text and Context*, New York 2016.
 Biber, Douglas, „A Typology of English Texts“, in: *Linguistics* 27 (1989), 3–43.
 Biber, Douglas, *Variation Across Speech and Writing*, Cambridge, UK, 1988.
 Biber, Douglas/Finegan, Edward, „Drift and the Evolution of English Style. A History of Three Genres“, in: *Language* 65 (1989), 487–517.

¹²⁰Tony Veale/Ekaterina Shutova/Beata Beigman Klebanov, „Metaphor. A Computational Perspective“, in: *Synthesis Lectures on Human Language Technologies* 9/1 (2016), 1–160.

¹²¹Rui Mao/ Chenghua Lin/ Frank Guerin, „End-to-end Sequential Metaphor Identification Inspired by Linguistic Theories“, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3888–98. <https://doi.org/10.18653/v1/P19-1378>. Siehe auch Alexander Friedrich/Chris Biemann, „Digitale Begriffsgeschichte? Methodologische Überlegungen und exemplarische Versuche am Beispiel moderner Netzsemantik“, in: *Forum Interdisziplinäre Begriffsgeschichte* 2/5 (2016), 78–96.

¹²²Erik-Lân Do Dinh/Iryna Gurevych, „Token-Level Metaphor Detection Using Neural Networks“, in: *Proceedings of the Fourth Workshop on Metaphor in NLP*, San Diego 2016. Siehe auch Jacobs/Kinder (Anm. 116).

¹²³Ich danke den beteiligten studentischen Hilfskräften Markus Paluch und Lisa Hannemann für lange Sitzungen zur Unzeit und vor allem für die dahinterstehende Motivation und Begeisterung! Dem *Campuslabor Digitalisierung und computergestützte Analyse in den Geistes- und Gesellschaftswissenschaften* der Universität Göttingen sei für die bereitgestellten Mittel gedankt. Im Februar 2017 wurden Teile der Forschungsarbeit am *Princeton Institute for Advanced Studies* durchgeführt. Hier danke ich herzlich Pat Geary von der School of Historical Studies und Maria Mercedes Tuyá von den *IAS Digital Scholarship Conversations*, wo ich erste Ergebnisse vortragen durfte.

¹²⁴Soweit nicht anders angegeben, wurden digitale Referenzen letztmalig am 1.9.2017 eingesehen und im Juli 2022 überprüft

- Biber, Douglas/Finegan, Edward, „Multi-Dimensional Analyses of Authors' Styles. Some Case Studies from the Eighteenth Century“, in: Don Ross/Dan Brink (Hg.), *Research in Humanities Computing*, Oxford 1994, 3–17.
- Bode, Katherine, „The Equivalence of ‚Close‘ and ‚Distant‘ Reading; or, Toward a New Object for Data-Rich Literary History“, in: *Modern Language Quarterly* 78/1 (2017), 77–106.
- Brooke-Rose, Christine, *A Grammar of Metaphor*, London 1958.
- Cohen, Jacob, *Statistical Power Analysis for the Behavioral Sciences*, Hillsdale, NJ 1988.
- Conrad, Susan/Biber, Douglas (Hg.), *Variation in English. Multi-Dimensional Studies*, London 2001.
- Deignan, Alice, *Metaphor and Corpus Linguistics*, Amsterdam/Philadelphia 2005.
- Detering, Heinrich/Sina, Kai, „Der deutschsprachige Roman 1900–1950“, in: Volker Meid (Hg.), *Geschichte des deutschsprachigen Romans*, Stuttgart 2013, 445–623.
- Do Dinh, Erik-Lân/Gurevych, Iryna, „Token-Level Metaphor Detection Using Neural Networks“, in: *Proceedings of the Fourth Workshop on Metaphor in NLP*, San Diego 2016, 28–33, <http://www.aclweb.org/anthology/W16-1104>.
- Dorst, Aletta G., „Personification in Discourse. Linguistic Forms, Conceptual Structures and Communicative Functions“, in: *Language and Literature* 20/2 (2011), 113–135.
- Dorst, Aletta G., *Metaphor in Fiction. Language, Thought and Communication*, Osterwijk 2011.
- Egbert, Jesse, „Style in Nineteenth Century Fiction. A Multi-Dimensional Analysis“, in: *Scientific Study of Literature* 2/2 (2012), 167–198.
- Eggs, Ekkehard, „Metapher“, in: Gert Ueding (Hg.), *Historisches Wörterbuch der Rhetorik*, Bd. 5, Tübingen 2000, 1109–1183.
- Evert, Stefan, „How Random is a Corpus? The Library Metaphor“, in: *Zeitschrift für Anglistik und Amerikanistik* 54/2 (2006), 177–190.
- Friedrich, Alexander/Biemann, Chris, „Digitale Begriffsgeschichte? Methodologische Überlegungen und exemplarische Versuche am Beispiel moderner Netzsemantik“, in: *Forum Interdisziplinäre Begriffsgeschichte* 2/5 (2016), 78–96.
- Gehring, Petra/Gurevych, Iryna, „Suchen Als Methode? Zu einigen Problemen digitaler Metapherndetektion“, in: *Journal Phänomenologie* 41 (2014), 99–109.
- Gibbs, Raymond W. Jr., „Process and Products in Making Sense of Tropes“, in: Andrew Ortony (Hg.), *Metaphor and Thought*, Cambridge, UK, 1993, 252–276.
- Gibbs, Raymond W. Jr., „Are Deliberate Metaphors Really Deliberate? A Question of Human Consciousness and Action“, in: *Metaphor and the Social World* 1 (2011), 26–52.
- Goatly, Andrew, *The Language of Metaphors*, London/New York 1997.
- Gries, Stefan Th., „Some Current Quantitative Problems in Corpus Linguistics and a Sketch of Some Solutions“, in: *Language and Linguistics* 16/1(2015), 93–117.
- Herrmann, J. Berenike, „Data-Driven Literary Studies“, (in Vorbereitung).
- Herrmann, J. Berenike, „Anschaulichkeit messen. Eine quantitative Metaphernanalyse deutschsprachiger Erzählanfänge zwischen 1880 und 1926“, in: Tilmann Köppe/Rüdiger Singer (Hg.), *Show, don't tell. Konzepte und Strategien narrativer Anschaulichkeit*, Bielefeld 2018.
- Herrmann, J. Berenike, „In a Test bed With Kafka. Introducing a Mixed-method Approach to Digital Stylistics“, in: Joris J. van Zundert/Sally Chambers/Marijn Koolen/Mike Kestemont/Catherine Jones (Hg.), *Digital Humanities Quarterly* 11/4 (2017), <http://digitalhumanities.org/dhq/vol/11/4/000341/000341.html>.
- Herrmann, J. Berenike, „Praktische Tagger-Kritik. Zur Evaluation des POS-Tagging des Deutschen Textarchivs“, in: *Tagungsband Fünfte Internationale Tagung Digital Humanities im deutschsprachigen Raum (DhD)*, Köln 2018, 287–290.
- Herrmann, J. Berenike, *Metaphor in Academic Discourse. Linguistic Forms, Conceptual Structures, Communicative Functions and Cognitive Representations*, Utrecht 2013.
- Herrmann, J. Berenike/van Dalen-Oskam, Karina/Schöch, Christof, „Revisiting Style, a Key Concept in Literary Studies“, in: *Journal of Literary Theory* 9/1 (2015), 25–52.

- Herrmann, J. Berenike, *Erzählanfangskorpus (EAK)*, Elektronische Ressource, <https://github.com/jberenike/EAK> (o.D.).
- Herrmann, J. Berenike/Woll, Karola/Dorst, Aletta G., „Linguistic Metaphor Identification in German“, in: Susan Nacey/Aletta G. Dorst/Tina Krennmayr/W. Gudrun Reijnierse (Hg.), *MIPVU in Multiple Languages*, Amsterdam 2019.
- Horton, William S., „Metaphor and Readers' Attributions of Intimacy“, in: *Memory & Cognition* 35/1 (2007), 87–94.
- Jacobs, Arthur M./Kinder, Annette, „What Makes a Metaphor Literary? Answers from Two Computational Studies“, in: *Metaphor and Symbol* 33/2 (2018), 85–100.
- Jakobson, Roman, „Zwei Seiten der Sprache und zwei Typen aphatischer Störungen“, in: Ders., *Aufsätze zur Linguistik und Poetik*, München 1979, 117–139.
- Jannidis, Fotis/Lauer, Gerhard/Martínez, Matías u.a., „Der Bedeutungsbe­griff in der Literaturwissenschaft. Eine historische und systematische Skizze“, in: Dies. (Hg.), *Regeln der Bedeutung. Zur Theorie der Bedeutung literarischer Texte*, Berlin 2003, 3–30.
- Jurish, Bryan, *Finite-state Canonicalization Techniques for Historical German*, PhD thesis, Potsdam 2012, <http://opus.kobv.de/ubp/volltexte/2012/5578/>.
- Kiesel, Helmuth, *Geschichte der deutschsprachigen Literatur 1918 bis 1933*, München 2017.
- Kilgarriff, Adam, „Language is Never, Ever, Ever, Random“, in: *Corpus Linguistics and Linguistic Theory* 1/2 (2005), 263–276.
- Koppel, Moshe/Schler, Jonathan/Argamon, Shlomo, „Computational Methods in Authorship Attribution“, in: *Journal of the American Society for Information Science and Technology* 60/1 (2008), 9–26.
- Lakoff, George, „The Contemporary Theory of Metaphor“, in: Andrew Ortony (Hg.), *Metaphor and Thought*, Cambridge, UK, 1993, 202–251.
- Lakoff, George/Johnson, Mark, *Metaphors we Live by*, Chicago 1980.
- Lakoff, George/Turner, Mark, *More Than Cool Reason. A Field Guide to Poetic Metaphor*, Chicago 1989.
- Lausberg, Heinrich, *Handbuch der literarischen Rhetorik. Eine Grundlegung der Literaturwissenschaft*, Stuttgart 1990.
- Mahlberg, Michaela, „Corpus Stylistics. Bridging the Gap between Linguistic and Literary Studies“, in: Michael Hoey/Michaela Mahlberg/Michael Stubbs u.a. (Hg.), *Texts, Discourse and Corpora*, London 2007, 219–246.
- Matlock, Teenie, „Metaphor, Simulation, and Fictive Motion“, in: Barbara Dancygier (Hg.), *The Cambridge Handbook of Cognitive Linguistics*, Cambridge, UK, 2017, 477–490.
- McHugh, Marry L., „Interrater Reliability. The Kappa Statistic“, in: *Biochemia Medica* (2012), 276–82.
- Michalke, Meik, „koRpus. An R Package for Text Analysis“ (Version 0.10–2, 2017), <https://reaktanz.de/?c=hacking&s=koRpus>.
- Nacey, Susan/Dorst, Aletta G./Krennmayr, Tina/Reijnierse, W. Gudrun (Hg.), *MIPVU in Multiple Languages*, Amsterdam 2019.
- Nöth, Winfried, *Handbuch der Semiotik*, Stuttgart/Weimar 2000.
- Pasma, Trijntje, *Metaphor and Register Variation. The Personalization of Dutch News Discourse*, Oisterwijk 2011.
- Perkuhn, Rainer/Keibel, Holger/Kupietz, Marc, *Korpuslinguistik*, Paderborn 2012.
- Pragglejaz Group, „MIP. A Method for Identifying Metaphorically Used Words in Discourse“, in: *Metaphor and Symbol* 22/1 (2007), 1–39.
- Rui Mao / Chenghua Lin / Frank Guerin, „End-to-end sequential metaphor identification inspired by linguistic theories“, in: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3888–98. <https://doi.org/10.18653/v1/P19-1378>.
- Saar, Ferdinand von, *Schloss Kostenitz. Novelle*, Heidelberg 1893.
- Semino, Elena/Steen, Gerard J., „Metaphor in Literature“, in: Raymond W. Gibbs Jr. (Hg.), *Cambridge Handbook of Metaphor and Thought*, Cambridge, UK, 2008, 232–246.

- Skirl, Helge, „Metaphorical Anaphors“, in: Monika Schwarz-Friesel/Manfred Consten/Mareile Knees (Hg.), *Anaphors in Text: Cognitive, Formal and Applied Approaches to Anaphoric Reference*, Amsterdam 2007, 103–120.
- Šklovskij, Viktor, „Die Kunst als Verfahren“, in: Jurij Striedter (Hg.), *Russischer Formalismus. Texte zur allgemeinen Literaturtheorie und zur Theorie der Prosa*, München 1994 (russ. 1916), 3–35.
- Sprengel, Peter, *Geschichte der deutschsprachigen Literatur 1900–1918: Von der Jahrhundertwende bis zum Ende des Ersten Weltkriegs*, München 2004.
- Steen, Gerard J., „Deliberate Metaphor Affords Conscious Metaphorical Cognition“, in: *Cognitive Semiotics* 5/1-2 (2009), 179–197.
- Steen, Gerard J., „The Paradox of Metaphor. Why we Need a Three-Dimensional Model for Metaphor“, in: *Metaphor and Symbol* 23/4 (2008), 213–41.
- Steen, Gerard J., *Finding Metaphor in Grammar and Usage*, Amsterdam/Philadelphia 2007.
- Steen, Gerard J., *Understanding Metaphor in Literature. An Empirical Approach*, London/New York 1994.
- Steen, Gerard J./Dorst, Aletta G./Herrmann, J. Berenike u.a., „Metaphor in Usage“, in: *Cognitive Linguistics* 21/4 (2010), 765–796.
- Steen, Gerard J./Dorst, Aletta G./Herrmann, J. Berenike u.a., <http://www.ota.ox.ac.uk/headers/2541.xml>, Oxford 2010.
- Steen, Gerard J./Dorst, Aletta G./Herrmann, J. Berenike u.a., *A Method for Linguistic Metaphor Identification. From MIP to MIPVU*, Amsterdam/Philadelphia 2010.
- Underwood, Ted, „A Genealogy of Distant Reading“, in: *Digital Humanities Quarterly* 11/2 (2017), <http://www.digitalhumanities.org/dhq/vol/11/2/000317/000317.html>.
- Veale, Tony/Ekaterina Shutova/Klebanov, Beata Beigman, „Metaphor. A Computational Perspective“, in: *Synthesis Lectures on Human Language Technologies* 9/1 (2016), 1–160.
- Vervaeke, John/Kennedy, John M., „Conceptual Metaphor and Abstract Thought“, in: *Metaphor and Symbol* 19/3 (2004), 213–32.
- Wagner, Karl, „Saar, Ferdinand Ludwig Adam von“, in: *Neue Deutsche Biographie (NDB)* 22 (2005), 315f.
- Winko, Simone, „Verstehen literarischer Texte versus literarisches Verstehen von Texten? Zur Relevanz kognitionspsychologischer Verstehensforschung für das hermeneutische Paradigma der Literaturwissenschaft“, in: *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* 69/1 (1995), 1–27.
- Wünsch, Marianne, *Realismus (1850–1890). Zugänge zu einer literarischen Epoche*, Kiel 2007.
- Zymner, Rüdiger, „Uneigentliche Bedeutung“, in: Fotis Jannidis/Gerhard Lauer/Matías Martínez u.a. (Hg.), *Regeln der Bedeutung. Zur Theorie der Bedeutung literarischer Texte*, Berlin 2003, 128–168.

Online-Ressourcen

- DTA-Leitlinien, <http://www.deutschestextarchiv.de/doku/leitlinien>.
Deutsches Wörterbuch der deutschen Sprache (DWDS), <https://www.dwds.de>.
Duden, <https://www.duden.de>.
Projekt Gutenberg-DE, <https://www.projekt-gutenberg.org/>

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Diskussionsbericht Sektion III.

Textanalyse

Jan Horstmann

In der Sektion *Textanalyse* wurde besonders hervorgehoben, dass strategische Argumentation von der Diskussion der methodischen Organisation und Umsetzung wissenschaftlicher Verfahren unterschieden werden sollten. Die *Digital Humanities* tendierten stark dazu, sich auf den zweiten Punkt zu konzentrieren. Auch formale, digitale Analysen in den Literaturwissenschaften agierten jedoch vor dem Hintergrund einer etablierten disziplinären Theorie und Praxis, zu der sie sich bewusst verhalten sollten. Für die *Digitale Literaturwissenschaft* sei es daher strategisch erforderlich, hypothesengeleitete und explorative Verfahren der digitalen Textanalyse zu integrieren, um als Methodenfeld größere Akzeptanz bei traditionellen Literaturwissenschaftlern zu finden. Betont wurde in diesem Zusammenhang auch die transformative methodologische Wirkung, die digitale Textanalysen für die Praktiken etablierter philologischer Disziplinen haben könnten.

Marcus Willand und Nils Reiter unterschieden in ihrer Vorlage literarische *Themen* (*Themes*) von *Topics* (die durch *Topic Modeling* generiert werden) und führten drei wörterbuchbasierte Ansätze und vier Einstellungen des *Topic Modeling* (figurespezifisch und übergreifend trainiert, mit je 20 und 100 Topics) durch. Anhand der Ergebnisse wurden die Methoden auf Grundlage von in der Literaturwissenschaft verbreiteten Annahmen als Postulate überprüft, womit der Anschluss digitaler Methoden an Konzepte und Wissen aus der Literaturwissenschaft gesucht werden sollte – ohne von einem Goldstandard auszugehen. Dass die einzelnen Tests nicht immer zufriedenstellend verliefen, wurde in der Diskussion als positiv aufgefasst. Die Neuartigkeit des Ansatzes wurde ebenfalls hervorgehoben: So behauptete Moretti in *Network Theory, Plot Analysis* (2011)

J. Horstmann (✉)
Westfälische Wilhelms-Universität Münster, Münster, Deutschland
E-Mail: jan.horstmann@uni-muenster.de

eine Plotanalyse durchzuführen, beschäftige sich jedoch eigentlich mit der Figurenkonstellation. Willand/Reiter hingegen untersuchten die Figurenkonfiguration und könnten deshalb auch eine tatsächliche Plotanalyse vornehmen.

Die Unterscheidung zwischen *Topics* und *Themen* fand breite Unterstützung, wobei der Unterschied zwischen den beiden Konzepten noch genauer herausgearbeitet werden sollte. So gebe es beispielsweise erfolgreiche *Topics*, die im engeren Sinne nicht thematisch seien, wohingegen *Themen* auch implizit vorhanden sein könnten. Eine Lemmatisierung oder ähnliche Methoden im Zuge des *Pre-Processings* beim *Topic Modeling* hätten auf der einen Seite eine erhöhte Vergleichbarkeit mit dem Wörterbuchansatz erzeugen können. Die Herangehensweise über das literaturwissenschaftliche Expertenwissen ermögliche es auf der anderen Seite, präzise nach Phänomenen zu suchen und unrelevante Daten schnell auszusortieren. Das kleine Set an Postulaten zeige bereits Unterschiede zwischen den angewandten Methoden. Dadurch ließen sich Empfehlungen ableiten, welche Methode für welche Frage die richtige sei. Dem Einwand, dass facheigene Postulate als gegeben modelliert würden und man daher auf literaturwissenschaftlicher Seite keinen Fortschritt konstatieren könne, wurde entgegengehalten, dass eine Toolentwicklung die Erzeugung von Daten ermögliche, die auch für die Entwicklung von neuen literaturwissenschaftlichen Fragestellungen genutzt werden könnten. Zudem könnten bisherige literaturwissenschaftliche Erkenntnisse verfeinert werden. Es herrschte Uneinigkeit darüber, ob ein methodisches Experiment aus dem Bereich *Distant Reading* eine literaturwissenschaftliche Erkenntnis fördern müsse, oder ob nicht die Reduktion der Methodendiversität eine kritische Reflexion – die stets Teil der disziplinären Diskussion sei – unterdrücke. Festgestellt wurde, dass der Beitrag von Willand/Reiter sich auf die Methodenentwicklung konzentriere. Die literaturwissenschaftlich orientierte Analyse und Interpretation der Ergebnisse sei erst der nächste Schritt und gefährde zu einem zu frühen Zeitpunkt die methodische Systematizität.

Im Zusammenhang mit der Korpuskonstitution wurde die Frage diskutiert, ob die sehr begrenzte Auswahl von kanonischen Texten nicht zu falschen Postulaten führen könne. Vor diesem Problem würde allerdings auch die klassische Literaturwissenschaft stehen, deren Erkenntnisse zwar häufig auf geringen Datenmengen basierten, deshalb jedoch nicht per se verworfen werden sollten. Auf Interesse stieß zudem die Genrefrage: Das untersuchte Korpus bestand ausschließlich aus Dramen; die Diskussionsteilnehmer waren sich uneinig darüber, ob z. B. Dramenenden (thematisch) klarer umrissen seien als die Enden von Romanen.

Positiv hervorgehoben wurde außerdem das Bemühen, traditionelle literaturwissenschaftliche Begriffe zu verwenden. Durch das äußerst komplexe *Black-Box-Verfahren* der einzelnen Methoden könne man klassischere Literaturwissenschaftler jedoch wieder verlieren. Es ergäben sich allerdings anschlussfähige literaturwissenschaftliche Fragestellungen. So sei es auf Grundlage der ermittelten Daten zukünftig z. B. möglich, unterschiedliche Figurentypen genauer zu analysieren. Die Operationalisierung selbst verlange jedoch nach einer stärkeren Validierung in zukünftigen Experimenten. Der Vorschlag, sie durch unstrittige synthetische Daten zu stützen, warf die Frage auf, welche Daten

tatsächlich unstrittig sein könnten. Synthetisch erzeugte Konfigurationen, die realistisch modelliert wurden, könnten bei Dramen hilfreich sein.

Maciej Maryl skizzierte in seiner Vorlage ein Forschungsprogramm, in dem der Wandel vom zentralistisch organisierten polnischen Literatursystem vor 1989 über eine Diversifizierung, d. h. Ausweitung der topographischen Herkunft und Kanonizität der Autoren, hin zu einer neuen – dieses Mal an marktwirtschaftlichen Prinzipien des kapitalistischen Systems orientierten – Zentralisierung im Zuge einer ‚Rückkehr‘ zu literarischen Zentren und kanonisierter Literatur am Ende des ersten Jahrzehnts des 21. Jahrhunderts untersucht wird. Dabei werden digitale Datenquellen wie Bibliografien, sowie literarische und kritische literarische Texte mit digitalen Methoden ausgewertet.

Als Problem wurde vor allem die Verfügbarkeit der Daten diskutiert. Die umfassende Bibliografie polnischer Literatur wurde zwar digitalisiert, die Zugriffsmöglichkeiten seien jedoch noch zu verbessern. Hier zeige sich das generelle Problem von Ansätzen in den *Digital Humanities*, mit gegenwärtigen Texten zu arbeiten. Im Gegensatz zu kontemporären *born-digital*-Texten sei die hier untersuchte prä-digitale Phase besonders schwierig. Pragmatische Gründe wie die digitale Verfügbarkeit und Rechtfreiheit führten daher häufig dazu, dass mit Texten älterer Jahrhunderte gearbeitet werde. Man solle Verlage und Herausgeber an dieser Diskussion teilhaben lassen, um Aspekte des Copyrights und Möglichkeiten der Forschung und des Datenaustauschs fundiert zu diskutieren.

Für das vorgestellte Forschungsprogramm wurde in der Diskussion eine Vielzahl von Vorschlägen gesammelt. So solle die Möglichkeit hervorgehoben werden, dass die Auswertung der Daten die erwarteten Ergebnisse nicht nur entweder belegen oder widerlegen werde, sondern auch noch etwas völlig Neues zeigen könnte. Zudem scheine eine Betrachtung anderer Formen der Zensur – wie durch finanzielle Förderung oder durch politisch motivierte Akteure – vielversprechend. Die Fokussierung lediglich einer Bibliografie und nicht der Inhalte der in dieser Bibliografie aufgeführten Werke selbst wurde als problematisch hervorgehoben: Die untersuchten Veränderungen sollten auch auf textinhaltlicher Ebene erforscht werden. Auch hierfür seien jedoch die digitale Verfügbarkeit und die urheberrechtliche Situation entscheidende Motive. Als mögliche Quellen für weitere Daten wurden in der Diskussion Schulbücher, *Wikipedia*-Einträge über Autoren oder urheberrechtfreie Gebrauchstexte wie beispielsweise das Feuilleton vorgeschlagen. Schwierigkeiten seien dabei, dass Schulbücher die Popularität von Autoren nicht verlässlich abbildeten und *Wikipedia*-Artikel häufig von den Autoren selbst geschrieben und daher ebenfalls kein verlässlicher Repräsentant von Popularität seien. Das Sammeln von Gebrauchstexten könnte schwierig werden, verspräche aber interessante Einsichten in die Diskussion populärer Texte.

Literaturwissenschaftliche Implikationen der Nutzung künstlich-neuronaler Netze in der Handhabarmachung umfangreicher Datenmassen beschrieb Christof Schöch in seiner Vorlage. Zugrunde lag die computerlinguistische semantische Hypothese, dass der Kontext von Wörtern in Verbindung mit ihrer Bedeutung stehe (*Distributional Hypothesis*; vgl. Harris 1954, Firth 1957). Im Verfahren des *word2vec* würden Vektoren, die Dokumenteigenschaften repräsentieren, stark in

ihrer Dimensionalität reduziert und dadurch handhabbar gemacht. Durch die neue Zwischenschicht der *Projection Layer* in den *Word Embedding Models (WEM)* werde es möglich, bei kleiner Dimensionalität Wörter verlässlich vorherzusagen. Das automatisierte Verfahren werde auf Grundlage der Oberflächendaten möglichst großer Korpora trainiert, sei nicht sprachspezifisch und modelliere Sprachkompetenz historisch genau. Ähnlichkeiten und Analogiebeziehungen zwischen Wörtern würden dadurch mathematisch greifbar. So könnten nicht nur rein syntaktische oder morologische, sondern vor allem auch semantische Ähnlichkeiten automatisch erkannt werden (z. B. „go“ und „went“, aber auch „apple“ und „mango“).

Auch der Variantenvergleich in der Editorik (vgl. *Sektion 2*) könne mit WEM spezifiziert werden, indem das Ausmaß der Veränderung quantifiziert werde. Das Verfahren stehe nicht in direkter Konkurrenz zu hermeneutischen Ansätzen, sondern erweitere als zusätzlicher Baustein auf plausible Art und Weise die klassische (menschliche) Sicht auf den Text. Durch eine Kombination unterschiedlicher methodischer Bausteine müsse das Potenzial des hermeneutischen Analyseprozesses vergrößert werden: Die Literaturwissenschaft stehe noch am Anfang dieser Entwicklung. In der Diskussion wurde darauf hingewiesen, dass klar kommuniziert werden müsse, dass es sich um eine relationale Semantik auf Wortebene handele, andere Ebenen wie bspw. die Satzebene jedoch noch nicht bedient werden könnten. Hier wurde ein engerer Dialog mit der Computerlinguistik nahegelegt, die viel zu diesem Thema forsche. Hingewiesen wurde auf die Notwendigkeit, sich der besonderen Art des relationalen Semantikverständnisses bewusst zu sein: So sei es in einem Korpus aus dem 19. Jahrhundert beispielsweise möglich, dass ‚Frau‘ und ‚Küche‘ im Netzwerk näher als ‚Frau‘ und ‚Mann‘ beieinander liegen. Die Einschränkung von WEM sei tatsächlich, dass es keine Rückbindung an Referenzialität gebe, sondern immer auf Grundlage von Relationen innerhalb des jeweiligen Korpus trainiert würde: Eine Verknüpfung z. B. mit Wörterbuchansätzen scheine daher vielversprechend. Auch ließen sich beispielsweise im Abgleich mit einem lyrischen Referenzkorpus punktuelle Abweichungen von der Sprachnorm (die konstitutiv für eine Gruppe von Lyrik sei) untersuchen.

Positiv hervorgehoben wurde in der Diskussion, dass der Forschungsprozess selbst in dem didaktisch sinnvoll strukturierten Beitrag sehr deutlich herausgestellt werde, wodurch ein Austausch zwischen *DH* und Literaturwissenschaft gezielter und besser möglich sei. Gleichzeitig sei es jedoch möglich, einzelnen Aspekten des Verfahrens eine Unterkomplexität vorzuwerfen, wenn z. B. ‚Prosa‘ und ‚Poesie‘ auch in Kontexten nah beieinander liegen, in denen dies nicht gewünscht werde. Als generelles Problem von Modellen, deren Verfahren man als menschlicher Nutzer nicht überprüfen könne, wurde die Schwierigkeit identifiziert, eine Übersicht über die eigenen Aktionen zu behalten und methodische Schief lagen zu entdecken. Die Grobheit der Methode sei es auch, die bestimmte (detailliertere) Fragestellungen ausschließe: So böte sich beispielsweise eine klassische Kookkurrenzanalyse besser an, um herauszufinden, wann ein Wort das erste Mal in einem neuen Kontext auftauche, da hierfür nicht die große statistische Masse, die WEM benötige, gebraucht werde. Die schiere Menge der benötigten

Daten (etwa 30 Mio. Wörter oder mehr wurden veranschlagt, um die Funktionalität von WEM zu gewährleisten) mache es schwierig bis unmöglich, die Methode auf einige historische Korpora (wie bspw. in der Mediävistik) anzuwenden, da dort eine solche Datenmenge nicht gegeben sei.

Im Zusammenhang mit dem differenztheoretischen Semantikmodell der WEM wurden schließlich Zusammenhänge und Unterschiede zum dekonstruktivistischen *Différance*-Begriff Derridas diskutiert. Es gebe im historischen System keinen archimedischen Punkt mehr; man könne die Methode als Rückkehr der DH in die Dekonstruktion auffassen. Andererseits böte die Saussure'sche Semantik bereits genügend Anhaltspunkte, um das hier skizzierte Verfahren zu greifen.

Die Arbeiten der *dlina*-Gruppe und ihre *Rabbit Holes* erfuhren in Peer Trilckes Beitrag eine Rekapitulation. Viele Projekte der *Digitalen Literaturwissenschaft* suchten nach literaturhistorischen Erkenntnissen, blieben dann jedoch bei Modellierungen und Operationalisierungen hängen. Dieser Prozess wurde demonstriert erstens anhand der *Small-Worlds-Studie*, zweitens anhand der *Beat-Charts-Studie*, welche die Rhythmik von Dramen anhand der Personalveränderung zwischen einzelnen Szenen mittels der Parameter *Drama Change Rate* bzw. *Segment Change Rate* beschreibt. So entstehe eine Dramentypologie auf Grundlage von Kennzahlen, die jedoch eine Eigendynamik entwickle: Es gehe nicht mehr allein um die Dramen, sondern auch um die Operationalisierung selbst (*Rabbit Holes*). Daraus ließen sich generelle Folgerungen schließen:

1. Digitale Begriffe seien zahlenmäßig geprägt und diese seien relational.
2. Es finde eine Anästhetisierung der untersuchten Objekte statt.
3. Daten und Zahlen würden häufig in Visualisierungen veranschaulicht, worin man eine Reästhetisierung sehen könne.
4. Digitale Projekte würden häufig eine Methode des Aufschiebs verfolgen: Man habe keine Abschlüsse, sondern befinde sich bewusst im Prozess.

Diskutiert wurde, mit Blick auf die *Small-Worlds-Studie*, zunächst die Sequenzialität der angewandten Methoden, die einen Ausschlussprozess initiieren, bei dem am Ende nur fünf Texte des gesamten Korpus übrig bleiben (*Faust*-Dramen und *Possen*), die man auch als relationale Extremtypen bezeichnen könnte. An dieser Stelle wurde eine Rückbindung an klassische literaturhistorische Diskussionen und ein Blick zurück in das gesamte Korpus gefordert. Auch wenn die Prozesshaftigkeit der dargestellten Verfahren erkannt – und von einigen Seiten auch kritisiert – wurde, betonten viele Diskussionsteilnehmer, dass Anknüpfungspunkte für die Literaturwissenschaft evident seien: Wo Irritationen im Gegenstand zu neuen Methoden führen, böten Irritationen im Modell immer auch einen Weg zurück zum Gegenstand. Diese Dynamik zwischen Modell und Gegenstand sei theoretisch sehr traditionell. Die virulente Methode des Unabgeschlossenen und der Gestus des Aufschiebs bei zurückgestellter Referenzialität ermöglichten eine gezielte Selbstreflexion der Literaturwissenschaft einerseits, zeigten jedoch andererseits auch gerade die Anschlussfähigkeit der demonstrierten Methoden: Ein umfassendes Mapping auf den literarischen Gegenstand in Form klassischer Publikationen und

eine Bereitschaft zur Falsifikation schienen am Platz. Die Differenz der diskutierten kommunizierten Netzwerke zu nicht-fiktionalen Netzwerken sollte ebenfalls näher betrachtet werden. Um über diesen Sondertypus von Netzwerken differenziertere Aussagen treffen zu können, brauche man allerdings sehr viel mehr Daten, die bspw. auch durch eine komparatistische Ausweitung auf andere Sprachen zu erlangen seien.

Zu überdenken sei außerdem die Rhetorik des Unvertrauten und Fremden, denn Literaturhistoriker seien nicht so verwundert über die Ergebnisse, wie im Beitrag behauptet werde. Die vier im Schlussteil diskutierten Punkte sollten geschärft werden, um auch die Unterschiede dieser Aspekte in Bezug auf die Literaturwissenschaft deutlich hervorzuheben. Problematisch sei hier beispielsweise der Begriff der ‚Reästhetisierung‘ im Zuge des – in den Literaturwissenschaften ungewöhnlichen – Erstellens von Diagrammen als Aussagen über die zuvor ‚anästhetisierten‘ Gegenstände. Alternativ wurde hier der Begriff der ‚Diagrammatisierung‘ stark gemacht. Die Diskussionsteilnehmer betonten jedoch, dass keine Analyse das literaturwissenschaftliche Ziel, alles über den untersuchten Gegenstand zu erfahren, je erreiche: Wo die quantitative Arbeit den ästhetischen Wert der Gegenstände vernachlässige, unterschlägen klassisch literaturwissenschaftliche Ansätze häufig die quantitative Dimension. Generell sollte die *Digitale Literaturwissenschaft* jedoch konzentrierter darüber diskutieren, ob das Digitale tatsächlich kategorial anders sei als das Nicht-Digitale, als das Postkoloniale, Dekonstruktive etc.

Thomas Weitin operationalisierte in seiner Vorlage Ähnlichkeiten über ein Textvergleichsmodell. Stilometrische Analysen wurden dabei als Netzwerk visualisiert und Unterscheidungskriterien sollten bis auf Wortebene bestimmt werden. Von der Analyse der durchschnittlichen Ähnlichkeit bahnte der Beitrag seinen Weg hin zur Analyse der maximalen Verschiedenheit. Als bemerkenswerte Punkte wurden in der Zusammenfassung genannt:

1. die Diagrammatisierung und Weiterverarbeitung der Daten mit dem *Simmelian-Backbone-Netzwerk*,
2. die stilometrische Ähnlichkeitsanalyse und -postulate sowie
3. die untersuchte literaturästhetische These über die stilistische Durchschnittlichkeit Heyses.

Beim letzten Punkt wurde empfohlen, die Diachronie des Textkorpus näher zu betrachten, um eine Durchschnittlichkeit a priori von einer Durchschnittlichkeit a posteriori (d. h. einer epigonalen Durchschnittlichkeit) unterscheiden zu können.

Das Plenum benannte als generelles Desiderat von *DH*-Ansätzen, das Begriffspaar von Ähnlichkeit und Unähnlichkeit, das vielen Methoden und Konzeptionen zugrunde liege, genauer zu diskutieren. Während der Beitrag für ein hypothesengeleitetes Vorgehen argumentiere, verfolge er selbst eher eine explorative Methode. Begründet wurde dies dadurch, dass eine historische Plausibilisierung der Fragestellung und ein hypothesengeleitetes Vorgehen immer der Operationalisierung und Analyse vorausgingen.

Als denkbare Anschlussfähigkeit der vorgestellten Untersuchung wurde die Möglichkeit genannt, Netzwerke zukünftig mit semantischen Werten zu verknüpfen und die einzelnen Werte zueinander in Bezug zu setzen, um z. B. die Frage beantworten zu können, wie viele Wörter den Unterschied bei einer Kategorisierung von Ähnlichkeit und Unähnlichkeit von Stil ausmachen. Empfohlen wurde, die Stabilität der Ergebnisse anhand unterschiedlicher Einstellungen der Rechenparameter zu überprüfen.

Das hier vorgestellte Verfahren zeige, wie stilometrische Analysen auch abseits der Autorschaftsattribute fruchtbar gemacht werden können. Der Distinktionsspielraum von Novellen sei gerade auf der Ebene des Stils (bzw. des ‚Tons‘) auszumachen, besonders im Epochenbruch von Romantik und Realismus. Die verwendeten Begriffe (wie z. B. der Stilbegriff) könnten dabei noch genauer literaturwissenschaftlich kontextualisiert werden.

Überlegt werden könne außerdem, ob ein Fokus auf die wenigen diskriminativen Ergebnisse der einzelnen ‚Ausreißer‘ nicht weniger fruchtbar sei, als stattdessen die Summe der Abweichungen zu betrachten – ein Vorgehen, das gerade im Zusammenhang mit der Stilfrage vielversprechend scheine. Das vorgestellte Verfahren impliziere ein Modell von Stil, das durch Durchschnittlichkeit und Abweichung definiert werde, wobei Stil durchaus auch anders umrissen werden könne.

Bei der Frage, welche Wörter dafür verantwortlich seien, dass einige Texte im Korpus als sehr unterschiedlich kategorisiert werden, könne das *Z-Score-Profil* als Autorschaftsprofil verstanden werden, da im Deltawert mehrere Dinge zusammengefasst würden. Zu fragen sei dabei jedoch, wie hilfreich die Beobachtung sei, dass das Profil eines Autors dadurch bestimmt werde, dass ein anderer Autor ein Wort besonders selten verwende.

Auch in diesem Beitrag wurde das neue Publikationsethos kontrovers diskutiert. Die literaturhistorische Perspektive sei bibliografisch sehr unterrepräsentiert, es müsse jedoch ein Bezug zu den einzelnen Forschungsfeldern hergestellt werden. Weitin wies darauf hin, dass er diese Perspektive an anderer Stelle bereits herausgearbeitet (siehe den Beitrag von Weitin) und sich hier auf die Korpusanalyse konzentriert habe. Zudem wurde konstatiert, dass sich der Habitus im Umgang mit Theorie ändere und jetzt stärker methodenorientiert sei, wodurch das Theorieverständnis nun auch ein mögliches Scheitern der jeweiligen Theorie beinhalte. Diese neue Konstellation münde in ein anderes Publikationsethos. So sei es sinnvoll, die Unterschiede zwischen den jeweiligen Forschungspraktiken zu suchen und den Gegenpol nicht auf die einfachste, sondern auf die komplexeste Weise zu konstruieren. Der Experimentcharakter vieler *DH*-Projekte ermögliche es, andere Ergebnisse als zuvor erwartet zu erlangen. Die Diskutierenden plädierten dafür, die *DH* solle zukünftig die Stärken beider Strömungen in sich vereinen.

Eine andere Form der Operationalisierung – nämlich eine, die sich in der Gestalt von Richtlinien an den menschlichen Annotator richtet – untersuchte Berenike Herrmann in ihrer Vorlage. Die manuell erstellten Annotationen wurden dabei zur quantitativen vergleichenden Auswertung genutzt. Während die

Annotation selbst eher einem *Close Reading* entspreche, komme durch die Analyse ein *Distant Reading* hinzu. Die kognitive Theorie der Metapher war in dem vorgelegten Ansatz das zugrundeliegende Konzept: Dabei wurden Eigenschaften einer Quelldomäne in eine Zieldomäne mit höherem Abstraktheitsgrad übertragen. Die Grundbedeutung eines Begriffs (Wörterbuchansatz) wurde mit der jeweiligen Kontextbedeutung verglichen, eine Abweichung wurde als Metapher annotiert. Dieses Verfahren zeitigte ein hohes *Inter-Annotator Agreement*.

Die Diskussionsteilnehmer hoben positiv hervor, dass der Aufsatz bemüht sei, das reiche Erbe der Metaphernforschung (siehe beispielsweise Lausberg) zu integrieren und dadurch sowohl ein literaturwissenschaftlich als auch ein eher technisch interessiertes Publikum anzusprechen. Als literaturwissenschaftlich anschlussfähig erwiesen sich bspw. Beobachtungen zur abnehmenden Satzlänge. Eine andere Studie zeige jedoch, dass zwischen 1850 und 1950 auch die Satzlänge in wissenschaftlichen Texten systematisch abnehme, es sich also um eine allgemeine Sprachentwicklung des bildungsbürgerlichen Sprechens und nicht nur um ein literarisches Phänomen handeln könne. Auch die konstatierte historische Abnahme der Menge an gebrauchten Metaphern lasse den Schluss zu, dass ebenfalls die Prozessleistung zur Verarbeitung von Metaphern abnehme. Der Ansatz sei gegenüber dieser phänomenologischen Komponente des Gegenstandsbereiches jedoch explizit agnostisch, da man keine Aussagen darüber treffen könne, wie Menschen mit den Metaphern umgingen. Eine Hypothesengenerierung sei indes auf Grundlage der Ergebnisse möglich. Diskutiert wurde zudem die Beobachtung, dass im skizzierten Verfahren nicht die abweichenden oder literarisch interessanten Metaphern im Vordergrund stünden. Trotz des weiten Metaphernbegriffs würden jedoch auch einige Metaphernspielarten aus der Analyse ausgeschlossen – bspw. etymologische Metaphern wie „zwiebeln“, da das Neuhochdeutsche als Referenz angelegt werde. Die pragmatische Entscheidung, bestimmte Formen von Metaphern auszuschließen und mit in Wörterbüchern festgelegten Bedeutungen zu arbeiten, wurde generell begrüßt, da so – neben dem *Peer-Review* innerhalb der Annotatorenteams – eine größere Unabhängigkeit vom jeweiligen Lektüreprozess gegeben sei. Eine Binnendifferenzierung in der Art der annotierten Metaphern fehle bislang noch. Zu Beginn möglichst breit zu annotieren, wurde als Methode jedoch begrüßt, da es zunächst darum ginge, eine große Menge an Daten zu erzeugen, an die spätere Automatisierungsversuche anschließen könnten. Außerdem wurde der Vorschlag gemacht, man solle die Metapher differenziell im Konkurrenzfeld zu den anderen Tropen (wie Metonymie, Ironie oder Emphase) untersuchen, um der Möglichkeit gerecht zu werden, dass ein Autor sehr unmetaphorisch, dafür aber stark ironisch schreiben könne.

In dem beschriebenen Experiment wurde mit Erzählanfängen gearbeitet, da gerade in den Anfängen von Erzählungen, in denen Text und Leser aufeinander treffen und die rhetorische Funktion, den Leser zum Weiterlesen zu animieren, erfüllt werden solle, interessante Textstellen vermutet wurden. Außerdem lasse die aufwendige Methode der manuellen Annotation keine großen Textmengen zu. Ob ein *Sampling* aus Volltexten ebenso gute Ergebnisse produziert hätte, sollte

überprüft werden. Als Alternative zum synchronen Wörterbuchansatz wurde über *word2vec* gesprochen, um Grund- und Kontextbedeutungen in zwei getrennten Korpora zuzuweisen. Wahrgenommen werden sollten zudem andere Ansätze, Metaphorizität zu modellieren, so z. B. der von Biemann u. a., die Jakobsons Poetizitätskonzept und eine distributionelle Semantik zugrunde legten.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.



Sektion IV. Schnittstellen der digitalen Literaturwissenschaft

Einführung

Thomas Stäcker

Eine Schnittstelle ist einerseits ein technisches Protokoll zur Nachrichtenübermittlung, andererseits bezeichnet sie im übertragenen Sinne die Übergangsstelle zwischen verschiedenen, in sich geschlossenen Systemen. Durch definierte Rahmenbedingungen bzw. Standards ermöglicht sie den Austausch von Daten, vermittelt aber auch ein Verständnis zwischen unabhängigen miteinander kommunizierenden Systemen. Schnittstellen bestehen auf der Ebene von Maschine zu Maschine oder von Mensch zu Maschine. Schnittstellen kann man allgemeiner aber auch als Metapher der Vermittlung verstehen. Eine standardisierte Schnittstelle ist somit nicht nur erforderlich, wenn man fremde Daten integrieren oder eigene Daten anderen zur Verfügung stellen möchte, ohne genau wissen zu müssen, wie das jeweils andere System im Inneren beschaffen ist. Der Begriff der Schnittstelle ist auch geeignet Eigenschaften zu beschreiben, die im Allgemeinen miteinander kommunizierenden Systemen zukommen. Damit ist das Thema der Schnittstelle auf verschiedenen Ebenen prädestiniert, die funktionalen und technisch-konzeptionellen Rahmenbedingungen der digitalen Literaturwissenschaften näher zu beschreiben.

Die Sektion IV hatte zum Ziel, unter dem Thema der Schnittstelle, das etwa in der Medientheorie intensiver diskutiert wird, vor allem die Frage der Standardisierung, der Interaktion und die Beziehung der digitalen Literaturwissenschaft zur digitalen Sammlung, vor allem, aber nicht nur dem Archiv und der Bibliothek, als den Reservoirs, die die digitale Literaturwissenschaft mit Daten und Dokumenten versorgen, zu adressieren. Dies war der Einsicht geschuldet, dass Digitalität einhergeht mit neuen Suchmöglichkeiten, mit Prozessen der Aggregation, der Interaktion oder auch Vernetzung, die ihrerseits aber nur dann

T. Stäcker (✉)
TU Darmstadt, Darmstadt, Deutschland
E-Mail: thomas.staecker@ulb.tu-darmstadt.de

wirksam werden, wenn sie in die Modellierungs- und Präsentationsverfahren der eigenen Forschung einfließen. Die Forderung nach Schnittstellenkonformität hat somit auf den verschiedenen Ebenen Auswirkungen auf den Forschungs- und Publikationsprozess und sollte in ihrer Tragweite und Virulenz diskutiert werden.

Diese Idee ließ sich leider nur teilweise umsetzen, was vermutlich darauf zurückzuführen ist, dass das Thema der Schnittstelle in der Literaturwissenschaft oder, konkreter, die Probleme der Standardisierung oder Interaktion in der digitalen Literaturwissenschaft in der täglichen Forschungsarbeit noch nicht in dem Maße im Vordergrund stehen, wie dies erwartet worden war, und sich bisher allenfalls als Ärgernis der Daten- und Dokumentbeschaffung und –nutzung aufdrängen. Insofern ist es auch überraschend, dass die digitale Literaturwissenschaft gegenüber den Infrastruktureinrichtungen, der Bibliothek und dem Archiv, nicht nachdrücklicher die zentrale Bereitstellung und Aggregation digitaler Daten und Dokumente für ihre Forschungen einfordert. Die frei verfügbaren Korpora sind, folgt man den üblichen Klagen, entweder von mäßiger Qualität oder angesichts der Gesamtproduktion immer zu schmal. Nur Textgrid, Gutenberg und das Deutsches Textarchiv werden bei den deutschen Korpora immer wieder genannt, reichen aber entweder vom Umfang her nicht aus oder genügen qualitativ den Anforderungen nicht. Andererseits sorgt die Literaturwissenschaft aber nur in begrenztem Umfang für die offene Bereitstellung ihrer Daten, geschweige denn, publiziert selbst regelhaft offen unter freien Lizenzen in standardisierten Formaten, obwohl sie, wie auch die Beiträge von Martus und Jannides gezeigt haben, sich selbst das Leben leichter machen würde, wenn die Zugänglichkeit der Daten und Dokumente, z. B. zur Reproduzierbarkeit der Ergebnisse, nicht durch rechtliche oder technische Hürden erschwert würden. All dies mag den Ausschlag gegeben haben, warum einerseits vergleichsweise wenige Beiträge des Symposiums sich selbst hier verortet, andererseits viele das Thema implizit oder explizit, gelegentlich auch unter anderen Überschriften, aufgegriffen haben. So kam in den vorangegangenen Sektionen z. B. die Problematik der Korpusbildung, deren Klassifikation und flexible fragenabhängige Zusammenstellung, die Vereinheitlichung der Annotation oder die standardisierte Struktur für Zeitschriften zur Sprache. Auch die Frage der Langzeitarchivierung, also des Reservoirs der Literaturwissenschaft, trat wiederholt hervor, etwa bei den verschwundenen Geniusdaten oder bei der Archivierung von Netzressourcen wie lebendigen Blogs, oder die Sorge um die langfristige Pflege von Oberflächen digitaler Editionen. Standardisierung schließt Fragen der Lemmatisierung oder des PoS-Tagging bzw. der Tokenisierung digitaler Editionen ein, wie im Beitrag von Brüning oder eben allgemein in der immer wieder sich aufdrängenden Frage nach brauchbaren Korpora für das Textmining. Dies ist vorauszuschicken, um die vielleicht hier und da möglicherweise nicht ganz passgenaue Platzierung eines Beitrags zu erklären.

In praktischer Hinsicht bestehen für die Literaturwissenschaft vielfältige Schnittstellenanforderungen, je nachdem, wo Aggregation oder Vernetzung nötig sind. Dazu kann ein einheitliches TEI Format gehören (z. B. TEI simple) oder aber die einheitliche Ansetzung von Entitäten (Personennamen, Orte, Zeitbegriffe, Gegenstände). Schnittstellenanforderungen ergeben sich auch mit Blick

auf den üblichen Kontext der Bereitstellung im Internet in der Beziehung der Dokumente untereinander (nach dem lateinischen Motto *liber librum aperit*), das auch in einigen Beiträgen zutage trat. Hier überraschte, dass das eigentlich naheliegende Thema der Hypertextualität und des *semantic web*, das für viele Fragen der digitalen Literaturwissenschaft Lösungen bereithält, bislang immer noch eine untergeordnete Rolle spielt. Effiziente Schnittstellen bieten Kataloge mit Metadaten, die nicht nur die Datenübernahme in eigene Publikationen erlauben und bibliographische Daten zur Verfügung stellen, sondern darüber hinaus auch eigene bibliometrische Auswertungen erlauben. Archiv und Bibliothek sind Repositorien für Daten und Texte und stellen sie idealerweise der digitalen Literaturwissenschaft in standardisierter und maschinenlesbarer Form zur Verfügung: Retrospektiv als digitale Edition oder digitalisierter und re-kodierter Text der schriftlichen Überlieferung, prospektiv als *born digital* oder elektronische Publikation. Sie bilden Infrastrukturen für geistes- und kulturwissenschaftliche Forschungsdaten, die das schriftliche kulturelle Erbe in eine maschinenlesbare, durchsuchbare und algorithmisch prozessierbare Form aufbereitet, nachhaltig sichert und der Forschung zur Verfügung stellt, die idealerweise ihrerseits Daten und Texte digital aufbereitet und annotiert, um sie dann in eine archivische oder bibliothekarische Infrastruktur zurückzuspielen. Wesentliche Bedingungen dieses auch als *research data life cycle* beschriebenen Prozesses sind neben technischen und informationswissenschaftlichen Anforderungen der digitalen Infrastruktur Nachhaltigkeitskonzepte, Standardbildung, hochwertige, qualitativ zuverlässige Daten und Texte und deren zentraler Nachweis über Metadaten, freie Zugänglichkeit idealerweise über Open Access Lizenzen und eine im DH – Konzept verankerte kooperative und inter- wo nicht transdisziplinäre Wissenschaftskultur.

In den Beiträgen zur Sektion spiegeln sich viele dieser intendierten Themen und Facetten der Schnittstelle. Während Stefan Dumont konkrete technische Schnittstellenanforderungen an den Austausch von Briefmetadaten vorstellt, geht es bei Karina van Dalen-Oskam um die Frage der Übersetzung literarischer Fiktion und damit eine innersprachliche Schnittstelle. Die Problematisierung des Werkbegriffs im Beitrag von Thomas Ernst führt zu einem Literaturbegriff, der soziale und prozessuale Aspekte, die im Netz immer auch als Schnittstellenthemen auftauchen, in die Literaturwissenschaft einführt. Querliegend zu klassischen Methoden ist die Messung von Umfängen eine Metrik, die zu überraschenden Einsichten in die Bewandnis von Bücherdicken führt und neue Zugänge zur Literatur eröffnet. Die Darstellung der Rekonstruktion des englischen Novellenschatzes von Allen Riddell wiederum greift mit dem Problem der Berechnung von Buchproduktion und Bestimmung von Verbreitungsszenarien auch Fragen der richtigen Schnittstellen zu diesen Daten auf. Die abschließende Darstellung von Thomas Stäcker entwickelt einen Begriff von Bibliothek, der sie in ihrer zentralen Schnittstellenfunktion für die Forschung sichtbar werden lässt.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Literaturwissenschaft und Bibliothek – Eine Beziehung im digitalen Wandel

Thomas Stäcker

1 Literaturwissenschaft und Bibliothek

Literaturwissenschaft und Bibliothek scheinen schon immer eng verbunden gewesen zu sein. Die Gründe dafür liegen auf der Hand. Die Literaturwissenschaft ist in besonderer Weise auf Literatur angewiesen. Die Bibliothek wiederum ist der Aufbewahrungsort für Literatur. Ohne Bibliothek also keine Literaturwissenschaft? Natürlich ist es nicht so einfach. Denn weder ist klar, von welcher ‚Bibliothek‘ die Rede ist, noch auch, von welcher ‚Literaturwissenschaft‘. Sieht man einmal von dieser Unschärfe ab, ist aber auch die Beziehung der Bibliothek zur Literaturwissenschaft nicht so eindeutig und eng, wie es den Anschein haben mag.¹ Zwar braucht die Literaturwissenschaft Literatur, aber braucht sie auch die Bibliothek? Man hat versucht, in Analogie zu den Naturwissenschaften die Bibliothek als *Laboratorium* zu verstehen,² in dem die Literaturwissenschaft gleichsam ihre Experimente durchführt. Allerdings deutet schon die Rede von ‚Experiment‘ die Sperrigkeit des Begriffs ‚Laboratorium‘ für einen Gebrauch innerhalb der Literaturwissenschaft an, denn die in den Literaturwissenschaften gängige

¹Andreas Brandtner 2015, „Bibliotheken als Laboratorien der Literaturwissenschaft? Innenansichten analoger, digitaler und hybrider Wissensräume“, in: Stefan Alkner/Achim Hölter (Hg.), *Literaturwissenschaft und Bibliotheken*, Göttingen 2015, 115–138, hier: 116.

²Jeffrey T. Schnapp/Matthew Balthes, *The Library Beyond the Book*, Cambridge, MA, u. a. 2014, 71 f.; Brandtner (Anm. 1), 116 f.; siehe beispielhaft die Profilbestimmung der *Herzog August Bibliothek*: <http://www.hab.de/de/home/wissenschaft/forschungsprofil-und-projekte/forschungsprofil.html> (letzter Aufruf 21.6.2018).

T. Stäcker (✉)
TU Darmstadt, Darmstadt, Deutschland
E-Mail: thomas.staecker@ulb.tu-darmstadt.de

hermeneutische Methode³ bedient sich in der Regel nicht Experimenten zum Erkenntnisgewinn.

Eine andere Richtung der Funktionsbestimmung der Bibliothek für die Literaturwissenschaft ist ihre Rolle als *Gedächtniseinrichtung* bzw. als *kulturelles Gedächtnis*.⁴ Die Bibliothek ist ein Archiv und Ort der Erinnerung, der Vergegenwärtigung der Literatur, deren Verständnis man sich im hermeneutischen Prozess aneignet. Als Archiv und Gedächtnis inhäriert ihr aber auch eine gleichsam mnemotechnische Ordnung, die das Suchen und Finden in ihren Beständen strukturiert. Diese bibliothekarische Ordnung versucht zwar einerseits, alle Fachspezifika abzubilden, andererseits aber auch temporäre Trends zugunsten einer stabilen Systematik zu glätten. Der Feind der Bibliothek ist die Zeit, die immer neue Ordnungssysteme generiert und die alten obsolet macht. Die Ordnung der Bibliothek ist nicht die Ordnung der Wissenschaft, auch wenn sie sich an ihr orientiert.⁵

Wegmann, der neben Werle zu den wenigen Literaturwissenschaftlern gehört, die sich des Konzeptes ‚Bibliothek‘ nicht als literarisches Motiv, sondern als Grundbegriff der Literaturwissenschaft⁶ angenommen haben, deutet auf dieses Missverhältnis, wenn er den Zugang der Literaturwissenschaft zur Bibliothek quer zur Bibliothek sucht und Konzepte wie *Serendipity* aktiviert, um das Unvordenkliche und gleichsam Dekonstruktivistische des Zugangs zur Bibliotheksliteratur zu charakterisieren. Die Heuristik der *Serendipity* lehrt, so Wegmann, dass es „keinen richtigen Weg“ gibt.⁷ Nicht die Ordnung gewährt den Zugang zur Literatur, sondern eine gegen sie widerständige und überbietende Praxis der, wie Wegmann es nennt: „Überinterpretation“.⁸

Der einflussreiche Wolfenbütteler Literaturwissenschaftler und Bibliothekar Paul Raabe wiederum sah in der Bibliothek eine *humane Anstalt*,⁹ die Bildungsaufgaben übernehme und Ort der intensiven Lektüre und schöpferischen geistigen Arbeit sei.¹⁰ Prägend war auch der von ihm wiederholt ins Feld geführte Begriff

³Zur Kritik der hermeneutischen Methode siehe Peter Rusterholz, „Zum Verhältnis von Hermeneutik und neueren anti-hermeneutischen Strömungen“, in: Heinz L. Arnold/Heinrich Detering, *Grundzüge der Literaturwissenschaft*, München 2001, 157–177.

⁴Maurice Halbwachs, *Das Gedächtnis und seine sozialen Bedingungen*, Berlin u. a. 1966; Jan Assmann, *Das kulturelle Gedächtnis. Schrift, Erinnerung und politische Identität in frühen Hochkulturen*, München 1992; Dirk Werle, „Literaturtheorie als Bibliothekstheorie“, in: Alkner/Hölter (Anm. 1), 13–26, hier: 22 f.

⁵Bernd Lorenz, *Systematische Aufstellung in Vergangenheit und Gegenwart*, Wiesbaden 2003, 45; vgl. Peter Blume, „Bibliothekarische Systematiken und Fachsystematik(en) der Philologien – eine vielschichtige Beziehung“, in: Alkner/Hölter (Anm. 1), 139–152.

⁶Nikolaus Wegmann, *Bücherlabyrinth. Suchen und Finden im alexandrinischen Zeitalter*, Köln u. a. 2000, 4.

⁷Ebd., 292.

⁸Ebd., 312.

⁹Paul Raabe, *Die Bibliothek als humane Anstalt betrachtet*, Stuttgart 1986; vgl. Werle (Anm. 4).

¹⁰Uwe Jochum, *Die Idole der Bibliothekare*, Würzburg 1995, 25 f.

des *Schatzhauses*.¹¹ Der hohe Ton trug nicht wenig zur Bewunderung bei, die diese Vorstellungen in breiten Bildungskreisen und auch in der Literaturwissenschaft weckten und ließen die Bibliothek nicht nur als einen Ort der wertvollen und klassischen Literatur, sondern auch *materialiter* der seltenen und kostbaren Bücher aufscheinen, der gleichsam als Gralsort des Geistes und, gebunden an die Materialität des schönen und wertvollen Buches, als kultureller Identitätsstifter verehrt wird. Diese mit den Vorstellungen der Bibliothek als Gedächtnisort verwandte Idee, sie als *Hüter der kulturellen Identität* zu begreifen, hat durchaus ihre Realität. Im negativen Sinne zeigt sie sich in kriegerischen Auseinandersetzungen, in denen militärisch eigentlich unsinnige Niederbrennungen oder Zerstörungen von Bibliotheken zu beobachten sind. Sie geschehen offenbar in der Absicht, feindliche Kulturen in ihrem Kern, ihrer kulturellen Identität, auszulöschen. Der antike Gemeinplatz dafür ist die Verbrennung der Bibliothek von Alexandria durch den Kalifen Omar, der neben dem *Koran* keine andere Literatur duldet und daher alles andere für überflüssig bzw. schädlich hielt.¹² In der Gegenwart kann man die Zerstörung der Bibliothek von Sarajewo durch serbische Truppen oder die von Timbuktu durch islamistische Milizionäre als traurige Beispiele dieser Logik nennen.

Auf der anderen Seite hat die Literaturwissenschaft auch immer wieder mit dem Kanon der schönen und wertvollen Bücher gekämpft und ihn bekämpft.¹³ Die Diskussion über kanonische Autorität, über Dichtung und Gebrauchstexte, über hohe und kanonische Literatur gegenüber der Alltagsliteratur reißt bis heute nicht ab. Die Bibliothek als *Ort der schönen Literatur*, die sich in ihrem Sammeln beschränkt, ist für diese Literaturwissenschaft zugleich ein Ort der kulturellen Vernachlässigung der Alltagsliteratur und des, mit Wegmann gesprochen, „Mülls“.¹⁴

Ein weiteres Motiv, das die Bibliothek mit Blick auf die Literaturwissenschaft kennzeichnet und das schon zur digitalen Bibliothek überleitet, ist die Sammlung als *Netzwerk*,¹⁵ verwandt mit literaturwissenschaftlichen Konzepten zur *Intertextualität*¹⁶ oder zum *Hypertext*.¹⁷ Intertextualität oder Hypertext war

¹¹ Siehe z. B. Raul Raabe, *Ein Schatzhaus voller Bücher. Die Herzog August Bibliothek in Wolfenbüttel*, Hannover ³1971.

¹² Diese Zerstörung ist übrigens mit demselben Argument auch von christlichen Autoren gutgeheißen worden und zeugt davon, dass diese Haltung religions- und kulturübergreifend ist, siehe z. B. Adam Contzen, *Politicorum Libri Decem*, Köln/Mainz 1621, lib. 2, cap. 4, § 20.

¹³ Heinz L. Arnold/Volker Sinemus (Hg.), *Grundzüge der Literatur- und Sprachwissenschaft*, Bd. 1: *Literaturwissenschaft*, München ⁶1980, 451 ff.; Simone Winko, „Literarische Wertung und Kanonbildung“, in: Arnold/Detering (Anm. 3), 585–600.

¹⁴ Wegmann (Anm. 6), 78.

¹⁵ Ebd., 226.

¹⁶ Matías Martínez, „Dialogizität, Intertextualität, Gedächtnis“, in Arnold/Detering (Anm. 3), 430–445, hier: 441 f., mit Hinweis auf Julia Kristeva und den Poststrukturalismus.

¹⁷ Wegmann (Anm. 6), 318; George P. Landow, *Hypertext 3.0. Critical Theory and New Media in an Era of Globalization*, Baltimore 2006; vgl. auch das Konzept eines *Docuversums* von Ted Nelson, auf das auch Wegmann hinweist.

schon unter anderen Vorzeichen und Begriffen in der Frühen Neuzeit ein Bestandteil der Gelehrtenkultur in der Gewissheit, dass *liber librum aperit*. Zahlreiche der damaligen der *historia litteraria* verpflichteten Gelehrten durchkämmten die Bibliothek in der Absicht, die Werke der jeweiligen Sammlungen zu ermitteln und miteinander in Beziehung zu setzen. Literaturwissenschaft, die es damals als solche freilich noch nicht gab (hilfsweise könnte man die *Literärgeschichte* der *res publica litteraria* dafür in Anspruch nehmen), tritt uns vor allem als propädeutische Wissenschaft der Ordnung des Wissens entgegen. Die Bibliothek ist demgegenüber die Büchermasse, die sich der Gelehrte einverleibt – allerdings weniger, um die Bibliothek im Sinne einer Bibliothekslehre „in Ordnung“ zu bringen, als um sich selbst zu einer „wandelnden Bibliothek“ zu einem Polyhistor umzuschaffen.¹⁸ Zugang zu den diversen Materien schafft denn auch weniger der Katalog oder die Struktur der Sammlung als der gelehrte polyhistorische Bibliothekar.¹⁹ Literaturwissenschaft und Bibliothek sind sich wohl nie wieder so nahe gewesen wie in der Person des frühneuzeitlichen Gelehrten.

Die überwiegende Wahrnehmung der Bibliothek seitens der Literaturwissenschaft ist heute allerdings diejenige einer *Gebrauchsbibliothek*, die nach ihrem jeweiligen Charakter (Nationalbibliothek, Universitätsbibliothek, Landesbibliothek, Forschungsbibliothek, Stadtbibliothek, Spezialbibliothek usw.) Dienstleistungen erbringt. Bibliothek ist die Institution, die weitgehend unbemerkt und geräuschlos Literatur und Quellen bereitstellt und einen Service am ‚Kunden‘ erbringt, dem man ‚Produkte‘ (vor allem Informationen) anbietet. Im innerbibliothekarischen Diskurs ist der Charakter der Bibliothek als reiner *Dienstleister* durchaus umstritten und wurde in der Debatte zur Dichotomie des philologisch-historischen und sozialwissenschaftlich-technologischen Paradigmas erbittert geführt.²⁰ Mit dem Abschneiden alter Zöpfe wurden in der Übernahme des sozialwissenschaftlich-technologischen Paradigmas zwar Wege freigeräumt für die Bewältigung der Informationsbedürfnisse der modernen Massenuniversität, doch verkümmerte, so die Gegner, zugleich die wissenschaftliche Seite der Bibliothekswissenschaft, der es zudem in Deutschland anders als in anderen Ländern nicht gelang, sich über den einen Standort an der Humboldt-Universität zu Berlin hinaus universitär zu etablieren. Dadurch gingen die althergebrachten inneren Verbindungen zur Literaturwissenschaft verloren oder konzentrierten sich auf wenige spezialisierte Bibliotheken. Die Literaturwissenschaft steht ihrerseits einer solchen Dienstleistungsbibliothek gleichgültig gegenüber. Ob die benötigte Literatur aus der Bibliothek, dem Buchladen, Antiquariat oder aus dem Netz

¹⁸Vgl. Daniel G. Morhof, *Polyhistor sive de notitia auctorum et rerum commentarii*, Lübeck 1688, lib. I, cap. III. de re bibliothecaria, 22, mit Verweis auf den von Longinus stammenden Begriff der ‚lebenden Bibliothek‘, <http://diglib.hab.de/drucke/ea-490/start.htm?image=00042> (letzter Aufruf 21.6.2018).

¹⁹Wegmann (Anm. 6), 129.

²⁰Vgl. Uwe Jochum, „Die Aufgabe des höheren Dienstes“, in: Ders. (Hg.), *Der Ort der Bücher. Festschrift für Joachim Stolzenburg zum 75. Geburtstag*, Konstanz 1996, 69–79.

kommt, wird zu einer Frage der Effizienz und Ökonomie, nicht der inneren Verbundenheit zur bereitstellenden Einrichtung. Die durchökonomisierte Bibliothek ist nur eines von vielen möglichen Verteilrelais und der Bibliothekar allenfalls ein literarischer *Mechanicus*. Die angeblich ‚geistlose‘ ökonomische oder nach einem Managementmodell²¹ geführte Dienstleistungsbibliothek ist oft mit der digitalen in einen Topf geworfen worden und rief in diesem Grundverständnis heftige ideologische Gegenreaktionen hervor.²² Doch hat, genau besehen, das eine mit dem anderen wenig zu tun. Das Digitale ist ebenso wenig der Untergang der Bibliothek wie der Literaturwissenschaft, noch auch zerstört es das „Gedächtnis“,²³ welcher Vorwurf geradezu eine kulturelle Konstante zu sein scheint, wenn man an die im Platonischen Dialog *Phaidros* überlieferte Kritik des König Thamus an Theuths Erfindung der Schrift denkt, die das Gedächtnis und damit die Fähigkeit zum Lernen und zur Weisheit schwächen soll. Der in der Bibliothekstheorie zentrale Begriff der *memoria* wird zwar mit der Digitalisierung zu einem externen Wissensspeicher, der technisch geordnet und abgerufen werden kann, doch unterscheidet sich dieser zumindest in diesem Punkt nicht von der traditionellen Bibliothek. Auch diese hat mechanische Hilfsmittel entwickelt (z. B. den Karteikasten oder eine systematische Buchaufstellung), um das Gedächtnis zu entlasten. Der Kern des bibliothekarischen Dienstleistungsparadigmas, demzufolge Wissen zur Ware oder zum Bereitstellungsproblem wird und darin zum bloß fakultativen oder potenziellen Wissen oder, mit griechischen Begriffen, von der Praxis zur Technik absinkt, ist daher keineswegs mit der Digitalisierung oder Digitalität in eins zu setzen. Die menschliche *memoria* bleibt im Sinne des kantischen Diktums *tantum scimus quantum memoria tenemus* Grundlage allen Wissens – gleichviel, ob sie digital unterstützt wird oder nicht, auch wenn literarische Phantasien wie Ted Williams' *Otherland* oder Konzepte des digitalen *Human Enhancements* sowie der totalen *Immersion*²⁴ solche Gewissheiten aufzuweichen scheinen. Umgekehrt zeigen letztlich die durch den Medienwandel evozierten neueren Ansätze ein produktives Aufnehmen des Dienstleistungsbegriffs, indem er sich von den konkreten Medien löst und auf den Wissenserwerb in den jeweiligen Zielgruppen abhebt: „The mission of librarians is to improve society through facilitating knowledge creation in their communities.“²⁵

Alle diese Seiten und Ausprägungen der Bibliothek als Laboratorium, Gedächtniseinrichtung, als Ort der kulturellen Identität, als humane Anstalt, Schatzhaus bzw. Ort der schönen Literatur, als Netzwerk oder Gebrauchsbibliothek

²¹ Jochum (Anm. 10), 25.

²² Vgl. insb. Uwe Jochum (Anm. 20), der sich dem vom Literaturwissenschaftler Roland Reuß initiierten Heidelberger Appel angeschlossen hat.

²³ Manfred Osten, *Das geraubte Gedächtnis. Digitale Systeme und die Zerstörung der Erinnerungskultur*, Frankfurt a. M. u. a., 2004, 78.

²⁴ Vincent Mosco, *The Digital Sublime. Myth, Power, and Cyberspace*, Cambridge, MA, 2005, 76.

²⁵ R. David Lankes, *The Atlas of New Librarianship*, Cambridge, MA, u. a. 2016, 13 ff.

bzw. Dienstleister treffen auf eine Literaturwissenschaft, die aus ihrer jeweiligen Konstitution heraus einen nicht minder vielfältigen praktischen oder theoretischen Gebrauch von der Bibliothek macht.

Neben Wegmann sieht vor allem Werle in der Bibliothek ein Forschungsfeld der Literaturwissenschaft und skizziert eine Literaturwissenschaft als Bibliothekswissenschaft.²⁶ Allerdings verleiht er der Literaturwissenschaft eine Speise ein, deren Zutaten nicht leicht verdaulich sind. Denn ‚die‘ Bibliothekswissenschaft, die sich in dem dreibändigen gleichnamigen Handbuch von Leyh u. a. ein Monument gesetzt hat,²⁷ ist längst in eine Vielzahl von sich teilweise überlappenden und miteinander konkurrierenden Disziplinen und Sparten zerfallen, die entweder eigene Fakultäten begründen konnten oder sich in anderen auflösten oder als propädeutische Wissenschaften reüssierten. Zu ihnen gehört die Buchwissenschaft, die Bibliotheksgeschichte, die Informationswissenschaft, die Bibliotheks- und Informationswissenschaft (*Library and Information Science*), die Bibliotheksverwaltung und -praxis, die Dokumentationswissenschaft, Hilfs- und Spezialwissenschaften wie Inkunabelkunde, Epigraphik, Paläographie, Kodikologie, Diplomatik bis hin zur Bibliographie und Editionswissenschaft. In den Komplex hineingezogen werden auch Felder wie Archivwissenschaft, Restaurierwissenschaften oder Museologie. Die Situation schildert Greetham mit Blick auf den Teilbereich der hier ebenfalls zuzurechnenden Textwissenschaften: „Textual studies is a discipline drowning in a sea of terms“.²⁸ Die Verbindungen zur Literaturwissenschaft liegen zwar auf der Hand, es ist aber im Einzelnen schwer abzugrenzen, welche Felder die Literaturwissenschaft auch für sich reklamieren kann. Will man jedoch eine Gemeinsamkeit in allen Feldern der Bibliothekswissenschaft identifizieren, so ist es sicher das Buch oder zumindest das Dokument als Gegenstand der Sammlung, Erschließung und Benutzung bzw. mit Blick auf die ökonomische, rechtliche und soziale Verwertung: seines Gebrauchs. Die Bibliothek bildet dafür das Gehäuse und den Referenzrahmen. Doch mit der Forderung, die Literaturwissenschaft als Bibliothekswissenschaft zu denken, lockert sich dieser traditionelle, von der Materialität des Mediums her gedachte Zusammenhang und betont stattdessen die ‚Literarität‘ der Bibliothek. Solche Ideen erinnern an ältere Gedanken der bibliothekarischen Zunft,²⁹ und der Frühneuzeitler Werle

²⁶Werle (Anm. 4).

²⁷Georg Leyh/Fritz Milkau/Karl Christ u. a. (Hg.), *Handbuch der Bibliothekswissenschaft*, 3 Bde., Wiesbaden 1931–1965. Der Begriff selbst geht auf Martin Schrettingers Lehrbuch der Bibliothekswissenschaft (1808–29) zurück.

²⁸David C. Greetham, *Textual scholarship. An introduction*, New York u. a. 1994, 1.

²⁹So formuliert der Grazer Bibliothekar Ferdinand Eichler: „Die Bibliothekswissenschaft ist die Erforschung der litterarischen Denkmäler mit Rücksicht auf die Voraussetzungen und die Art ihrer Entstehung, Verbreitung und Benützung!“ (Ders., *Begriff und Aufgabe der Bibliothekswissenschaft. Vortrag gehalten im Oesterreichischen Verein für Bibliothekswesen in Wien am 30. Mai 1896*, Leipzig 1896, 17). Die Bibliothekswissenschaft dürfe ihre Aufmerksamkeit nicht nur auf die Einrichtung und Verwaltung von Bibliotheken richten: „Ihr Ziel ist, wie das jeder anderen Wissenschaft, die Wahrheit zu erkennen, das ist in diesem Falle die Wahrheit von dem Werte

knüpft hier vielleicht nicht von ungefähr auch an Ideen der *Literärwissenschaft* an, deren Bemühungen sich nicht so sehr auf die Verzeichnung und den Erwerb von Literatur, sondern zum Zwecke des Erwerbs von Gelehrsamkeit auf deren intellektuelle Aneignung richten, eben um eine „wandelnde Bibliothek“ (s. o.) zu werden, was hier mehr als nur eine Metapher ist.

Bibliothek und Literaturwissenschaft stehen, wie sich hier zumindest skizzenhaft zeigt, in einer sehr komplexen Beziehung zueinander, die man vergegenwärtigen muss, wenn es um die Beziehung der Literaturwissenschaft zur Bibliothek unter digitalen Bedingungen geht. Denn die *digitale Literaturwissenschaft* modifiziert zwar die Anliegen der Literaturwissenschaft an die Bibliothek, ändert sie aber nicht grundsätzlich. Umgekehrt verändert die *Digitale Bibliothek* die Literaturwissenschaft, ohne aber selbst Charakteristika preiszugeben, die sie als Bibliothek kennzeichnen. Digitalität muss daher auch mit Blick auf die Auswirkungen auf die Bibliothek als Laboratorium, als Gedächtniseinrichtung, als Ort der kulturellen Identität, als Schatzhaus, als Netzwerk und Dienstleister oder auch konzeptioneller Gegenstand der Literaturwissenschaft bedacht und hinterfragt werden.

2 Text und Daten

Mit dem Übergang zum Digitalen ändert sich die Form des Gegenstandes für die Literaturwissenschaft³⁰ bzw. literaturwissenschaftlich geprägte Bibliothekswissenschaft. Man mag zunächst einwenden, dass sich mit dem Medienwechsel lediglich das Kommunikationsmittel ändere und dass der Inhalt der Literatur davon unberührt bliebe (ihr ‚Sinn‘ oder ihre ‚Bedeutung‘). Das trifft zwar einerseits zu, Goethes *Faust* ist immer noch Goethes *Faust*, ganz gleich, ob er auf Papier oder digital publiziert wird, doch wissen wir spätestens seit McLuhans Untersuchungen³¹ um den Einfluss des Mediums auf die Fragestellungen und Gegenstände der jeweiligen Wissenschaft.³²

literarischer Denkmäler.“ (Ders., *Bibliothekswissenschaft als Wertwissenschaft. Bibliothekspolitik als Weltpolitik*, Graz/Leipzig 1923, 7).

³⁰Mirco Limpinsel, „Was bedeutet die Digitalisierung für den Gegenstand der Literaturwissenschaft?“, in: *Zeitschrift für digitale Geisteswissenschaften* 2016, https://doi.org/10.17175/2016_009.

³¹Marshall McLuhan, *Understanding Media*, London u. a. 1964.

³²Vgl. Limpinsel (Anm. 30): „[...] besteht Anlass zur Vermutung, dass philologische Fragestellungen ihrerseits ein Effekt von Medienkulturen sind, dass, mit anderen Worten, die Ausübung von Philologie eine Funktion hat, die für Bücher sinnvoll war, für digital vorliegende Texte aber möglicherweise nicht.“ Ein gutes Beispiel ist auch die genetische Edition [siehe Martens]. Sie verliert weitgehend ihren Sinn, wenn der Autor nur noch an dem einen digitalen Text schreibt und frühere Versionen überschreibt. Die Arbeit des Autors, der vormals noch Manuskripte *sensu stricto* angefertigt hat, ist in der typischen Situation des Verfassens von Texten am Computer nicht mehr sichtbar und kann allenfalls in Differenzanalysen verschiedener Dateien nachvollzogen werden.

Im Digitalen tritt uns Literatur in zweierlei Form entgegen: als (retro-) digitalisierte oder als digitale Literatur.³³ Während erstere unter den Bedingungen des Gutenbergzeitalters (oder davor) verfasst wurde und ins digitale Medium überführt wird, steht letztere unter digitalen Bedingungen und wird unter bewusstem Einsatz digitaler Mittel erstellt, so genannte *born-digitals*. Literatur, die die traditionelle Form im Digitalen simuliert (die digitale Inkunabel, typischerweise ein PDF) ist als Übergangsphänomen zu betrachten.³⁴ Beide Verfahren zielen darauf, das im Digitalen liegende Potenzial für sich zu nutzen.³⁵ Die (Retro-) Digitalisierung verhält sich wie seinerzeit die Übertragung von Handschriften in Drucke in Zeiten des Humanismus bzw. am Anfang des Gutenbergzeitalters und kann insofern als Remediativierung³⁶ oder auch Recodierung³⁷ verstanden werden. Die Änderungen sind weitreichend, auch wenn es auf den ersten Blick nicht so scheinen mag. Für die Bibliothek bedeutet die Wende zum Digitalen einen massiven Umbruch, den man durchaus als Paradigmenwechsel bezeichnen kann. Versuchsweise könnte man ihn dadurch charakterisieren, dass sich die Bibliothek der Bücher zu einer Bibliothek der Texte wandelt.³⁸ Der hier verwendete Begriff des ‚Textes‘ ist indes auslegungsbedürftig.³⁹ In der Literaturwissenschaft ist ‚Text‘ meist, aber nicht immer schriftgebunden. So versteht ihn auch die Textlinguistik, die sich über den Begriff definiert.⁴⁰ ‚Text‘ steht als umfassender Begriff in Konkurrenz und Beziehung zu anderen Begriffen, z. B. den des ‚Werkes‘ oder des

³³ Zum Unterschied Patrick Sahle, *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, 3 Bde., Norderstedt 2013, hier: Bd. 2, 152; vgl. auch Roger T. Pédaque, *Le document à la lumière du numérique. Présenté par Jean-Michel Salain*, Caen 2006.

³⁴ Problematisierend zu PDF als Stabilitätsgarant siehe Pédaque (Anm. 33).

³⁵ Beide Bereiche konvergieren in ihrer Absicht, siehe ebd.

³⁶ Jay D. Bolter/Richard Grusin, *Remediation. Understanding new media*, Cambridge, MA, [u. a.] 2000.

³⁷ Sahle (Anm. 33), Bd. 3.

³⁸ Thomas Stäcker, „Vom Buch zum Text. Sammeln, Erschließen und Benutzen im digitalen Zeitalter“, in: *Buch – Bibliothek – Region. Wolfgang Schmitz zum 65. Geburtstag*, Wiesbaden 2014, 353–356.

³⁹ Aus der reichen Literatur zum Textbegriff im digitalen Medium vgl. z. B. Paul Eggert, „Text-encoding, Theories of the Text, and the ‚Work-Site‘“, in: *Literary and Linguistic Computing 20/4 (2005)*, 425–435; Landow (Anm. 17); Pédaque (Anm. 33) oder Paul Caton, „On the term ‚text‘ in digital Humanities“, in: *Literary and Linguistic Computing 28 (2013)*, 209–220. Allgemein zur Theorie der Texte siehe den Reader *Texte zur Theorie der Texte*, Stuttgart 2005 (hg. von Stephan Kammer/Roger Lüdeke). Ein umfassendes Textmodell liefert Sahle (Anm. 33), Bd. 3, 46, indem er sechs Textbegriffe unterscheidet. Er unternimmt aber keinen Versuch, den hier versuchten digitalen Begriff von ‚Text‘ in sein Textmodell einzugliedern.

⁴⁰ Monika Schwarz-Friesel/Manfred Consten, *Einführung in die Textlinguistik*, Darmstadt 2014.

‚Dokuments‘, und wechselt wie diese je nach Anwendungsbereich, Fragestellung und Kontext seine Bedeutung.⁴¹ Trotz dieser Unschärfen lassen sich im Digitalen Merkmale festhalten, die es erlauben, den Text vom Dokument oder Werk zu unterscheiden. So diskutiert Pédaque⁴² den Begriff des ‚Dokuments‘ nach den Kriterien *forme*, *signe* und *medium*.⁴³ Ich möchte den Begriff des ‚Dokuments‘ jedoch primär unter dem Begriff *forme* betrachten, also primär vom Substrat bzw. von der Form her denken, sei es Papier oder eine Datei. Ohne den Gedanken eines Substrates gibt es kein Dokument. Text gehört eher der Dimension von *signe* und *medium* an. In diesem Sinne würde ich auch Barthes' bekanntes Dictum verstehen wollen, dass Text in der Sprache sei und das Werk in der Hand liege, wobei ich für ‚Werk‘ hier den Begriff des ‚Dokuments‘ bevorzugen würde. Man kann, wie Barthes ausführt, den Text daher nicht in die Regale von Bibliotheken stellen, wohl aber die papiersubstratgebundenen Dokumente (Bücher). Wichtig für den Begriff des ‚Textes‘ ist hier, dass er anders als das Dokument oder das Werk im Sinne Barthes'⁴⁴ schon im vordigitalen Zeitalter von seinen materiellen Grundlagen gelöst erscheint und einen sprachlichen bzw. schriftlichen Charakter trägt.⁴⁵

Gleichwohl, selbst wenn im vordigitalen Zeitalter Texte im Prinzip ‚transitorisch‘ waren, wird diese Eigenschaft erst im Digitalen virulent – nämlich darin, dass der Träger gegenüber dem Text weitgehend gleichgültig und für seine Gestaltung und Subsistenz unerheblich wird. Weder das Aussehen der Schrift noch auch deren Subsistenz haftet noch am Träger.⁴⁶

Die Loslösung vom Träger (Papier, Pergament etc.) vollzieht sich über die Binarisierung. Sie liegt allen elektrischen Ladezuständen oder auch Speicherkonzepten (magnetisch, optisch, elektronisch) zugrunde. ‚Binarisierung‘ als Synonym der ‚Digitalisierung‘ bedeutet, dass Gegenstände in einen binären Code von Nullen und Einsen übersetzt werden. Das Dokument nimmt diesen Code als Ladungszustand auf und wird zur Datei oder zu einem Zusammenhang

⁴¹ Caton (Anm. 39).

⁴² Ich bezeichne den Kollektivautor Pédaque bequemerweise mit dem selbst gegebenen Namen.

⁴³ Pédaque (Anm. 33).

⁴⁴ Der moderne bibliothekarische Gebrauch, der dem Modell der *Functional Requirements for Bibliographic Records (FRBR)* folgt, verwendet den Begriff ‚Werk‘ als Abstraktion, als geistige Entität, die die Einheit von *Expression*, *Manifestation* und *Item* gewährleistet.

⁴⁵ Roland Barthes, „De l'œuvre au Texte“, in: *Revue d'esthétique* 3 (1971), 225–232, hier: 226; Barthes (ebd., 227): „Il s'ensuit que le Texte ne peut s'arrêter (par exemple à un rayon de bibliothèque); son mouvement constitutif est la traverse (il peut notamment traverser l'œuvre, plusieurs œuvres)“. Siehe zur Problematik der Schrift auch die grundsätzlichen Überlegungen von Jacques Derrida in *Schrift und Differenz*, Frankfurt a. M. 1972 [frz. 1967].

⁴⁶ Deswegen ist auch die Diskussion um die Langlebigkeit von DVDs oder Festplatten überflüssig, weil sie nichts zur Textkonstitution beitragen. Durch einen einfachen Kopierprozess wird der Text verlustfrei von einem Medium auf das nächste transportiert. Zur eigenen Materialität des digitalen Buches siehe Constanze Baum, „Im Hyperfluss“, in: Dies./Ulrike Gleixner/Jörn Münker u. a. (Hg.), *Biographien des Buches*, Göttingen 2017, 326–348, hier insb.: 337 ff.

von Dateien.⁴⁷ Der binarisierte Text wiederum besteht aus sogenannten *Code Points* (Unicode⁴⁸), denen Schriftzeichen zugeordnet sind. Das System ist – wie schon die Typographie – gegenüber der handschriftlichen Form reduktionistisch. Viele Informationen der Handschrift oder der Darstellung lassen sich auf der Code-Ebene nicht oder nur unvollständig wiedergeben bzw. nicht der Menge der Textzeichen zuordnen. Sie sind dadurch entweder kein Text oder betreffen seine vom Code zu trennende Darstellung: das Layout (z. B. Unterstreichungen, figürliche Initialen etc.). Grundlage des digitalen Textes in diesem Sinne ist daher der *Code*, genauer gesagt: eine Menge von theoretisch 1.114.112 Code-Punkten (65.536 Zeichen * 17 mögliche *Planes* in Unicode), die – mit Glyphen verbunden – als Text repräsentiert werden.⁴⁹ Eine Bibliothek der digitalen Texte ist damit zunächst eine Bibliothek von maschinenlesbaren *Code Points*, die über einen Textrenderingprozess oder *Text Processes*, wie es im Unicode-Standard heißt, zu ‚Text‘ generiert werden. Der Text ist daher Resultat eines Algorithmus,⁵⁰ der Grunddaten (*Codes*) mit Darstellungsinformationen verbindet. ‚Text‘ in diesem Sinne ist aber nicht nur ein mit einem Font gerendertes lesbares Zeichen, sondern als Code ohne diesen Font maschinenlesbar, Voraussetzung für jede digitale Anwendung. Als solche nennt sie das W3C auch *Character Data*. Sie bilden die digitalen Bausteine für Wörter und Sätze, sind aber auch Grundlage für die Zuordnung von Eigenschaften, die für das Verständnis des Textes unverzichtbar sind.

Dem Text kommen indes spezifische Eigenschaften zu, die durch die Code Points allein nicht erfasst werden. In der Literaturwissenschaft spricht man von ‚Textstrukturen‘, die nach der in der Disziplin typischen Unterscheidung phonetisch, syntaktisch, semantisch oder pragmatisch analysiert werden können. Zu den Textstrukturen gehören auch visuelle Informationen, wie Absätze oder abgesetzte Fußnoten, Seitenzahlen u. ä. Die westliche Schrift- und Druckkultur hat Gewohnheiten des *mise en page* entwickelt, die Bedeutungen an ein bestimmtes Aussehen knüpfen, auch wenn sie innerhalb des europäischen Kulturkreises variieren (Endnoten oder Fußnoten, Literaturverzeichnisse und Inhaltsverzeichnisse vorn oder hinten).⁵¹ All dies muss ebenfalls entweder aus einer analogen Grundform in eine digitale Form ‚übersetzt‘, re-codiert, oder aber im digitalen Text explizit gemacht werden. Dafür bieten sich verschiedene Verfahren an. Von

⁴⁷Das Konzept der ‚Datei‘ (‚file‘) ist bisher in der Textwissenschaft nicht gleichrangig zu anderen wie Buch oder Dokument diskutiert worden, wiewohl sie eine zentrale Rolle bei allen digitalen Medien spielt. Eine Kritik der Formate (.txt,.docx,.odt,.pdf,.xml etc.) findet meist außerhalb der engeren Textwissenschaft statt.

⁴⁸<http://www.unicode.org> (letzter Aufruf 21.6.2018); gängig ist in westlichen Breiten UTF-8.

⁴⁹The Unicode® Standard. Version 12.0 – Core Specification, hier: „Architectural Context“, 10, <https://www.unicode.org/versions/Unicode12.0.0/UnicodeStandard-12.0.pdf> (letzter Aufruf 8.6.2019).

⁵⁰Ebd., „Conformance Requirements“, 79.

⁵¹Mit diesem Bereich beschäftigt sich u. a. auch die Schriftlinguistik.

den Elementen, den *Character Data*, her gedacht, ist ‚Text‘ in der Regel eine sequenzielle Abfolge von Einzelzeichen, diese bilden Wörter, diese wiederum Sätze, diese ggf. ein abgeschlossenes Werk, dieses ein Netzwerk. Die Abfolge der Zeichen, Worte und Sätze ist für das Verständnis essenziell. Allerdings kann die Richtung variieren (Leserichtung links-rechts, rechts-links, oben-unten, Bustrophedon). Was genau ein Wort oder Grundelement der Sprache ist, muss definiert werden (etwa durch Tokenisierung).⁵² In der Regel dienen heute Leerzeichen bzw. Whitespaces oder Satzzeichen dazu, Wort- und Satzgrenzen zu bestimmen, doch ist das der Schrift nicht unbedingt immanent, wie z. B. frühe Unzialschriften lehren. Die Worttrennungen ergeben sich durch die laut lesende, die Worte artikulierende Stimme; Worte selbst können, wie im Deutschen typisch, in Teile zerfallen. Im Digitalen bräuchte man dafür einen eindeutigen Code, denn die dafür im Layout typischerweise verwendeten Zeichen wie das Leerzeichen oder der Punkt sind mehrdeutig. Um Texte dieserart aufzubereiten, ist es daher erforderlich, die grammatischen und satzstrukturellen Einheiten unzweifelhaft zu qualifizieren (z. B. durch *POS Tagging* oder strukturelle Auszeichnungsmerkmale bzw. *Markup* bzw., allgemeiner formuliert, Annotation).

Die Frage, die sich stellt, ist, was man als Textstandard im Digitalen zugrunde legt, denn offensichtlich sind Kodierungen dieser Art variabel, technikgebunden und disziplinenabhängig. Während sich bei den *Character Data* Unicode als Standard etabliert hat, folgt die Kodierung von Textstrukturen variablen Textmodellen und Kodierungsverfahren, deren Standardisierung allenfalls in Teilen (vgl. z. B. die Aktivitäten der TEI⁵³) erfolgt ist und noch der weiteren Konsolidierung bedarf. Man darf allerdings nicht vergessen, dass es auch im Gutenbergzeitalter Jahrzehnte gedauert hat, bis ein Konsens erreicht werden konnte, was z. B. Fettdruck ‚bedeutet‘, was Überschriften sind, wie man Fußnoten über ihre Positionierung versteht oder Texte gliedert. Bis heute gibt es Unklarheiten, manchmal auch bewusst als solche im Schreiben eingesetzt. Typisch ist die Kursive, die, wenn sie nicht für Zitate verwendet wird, vieles bedeuten kann: Emphase, ironische Bemerkung, fremdsprachlicher Ausdruck, Name etc. Layout hat daher eine bedeutungstragende Funktion, die bei der digitalen Übersetzung explizit gemacht werden muss, um den Bedeutungsumfang des Textes vollständig wiederzugeben. Das dafür verwendete Markup⁵⁴ bildet die Grundlage, um scriptgesteuert eine am Bildschirm oder im Druck lesbare Form zu erzeugen. Wie die *Code Points* für den *Plain Text*, so werden die Strukturen für das *mise en page* genutzt. Mit der Einführung von SGML bzw. heute XML hat es dabei

⁵²Vgl. auch Jerome McGann, *A New Republic of Letters. Memory and Scholarship in the Age of Digital Reproduction*, Cambridge, MA/London 2014, 91 f., mit Verweis auf Susan Hockeys Aussage: „There is no obvious unit of language“.

⁵³<http://www.tei-c.org> (letzter Aufruf 21.6.2018).

⁵⁴Übersicht bei Sahle (Anm. 33), Bd. 3, 99 ff.; vgl. James H. Coombs/Allen H. Renear/Steven J. DeRose, „Markup Systems and the Future of Scholarly Text Processing“, in: *Communications of the ACM* 30 (1987), 933–947, <https://doi.org/10.1145/32206.32209>.

eine wichtige Entwicklung gegeben. Während frühere Systeme durch sogenanntes *Procedural Markup* das Layout bzw. die Seitendarstellung unmittelbar kodierten (sodass es nur von einer bestimmten Software dargestellt werden konnte), löst XML das Markup von seiner festen Verbindung mit einer Software und trennt so nicht nur Struktur und Inhalt, sondern ermöglicht auch durch eine neue Schicht des *Semantic Markup* die explizite Zuordnung von ‚Bedeutung‘ zu Textteilen und wiederum über Scripte der XML-Familie (XSLT, XQuery) ein flexibles *Layouting*, wobei ‚Layout‘ eine besondere Form der Transformation ist, denn semantische Elemente können nicht nur mit Layout versehen, sondern auch in neue Formen überführt werden (z. B. Umwandlung in HTML, EPUB oder SVG). Der darin liegende Textbegriff nach der XML-Spezifikation ist eine Kombination von Inhalt und Struktur.⁵⁵ In der W3C-Definition heißt es: „Text consists of intermingled character data and markup“, wobei für „character data“ gilt: „All text that is not markup constitutes the character data of the document“.⁵⁶ „Character data“ (Unicode-Text) und „markup“ bilden so zusammen den Text, der sich in einem „document“ bzw. Dateien materialisiert.

Natürlich muss der Computer wissen, was das jeweilige Markup bedeutet. Da XML keine Regeln über die Benennung der Tags mitbringt, muss das Markup durch sogenannte Schemadateien (DTD, W3C-Schema, Relax-NG) festgelegt werden, um nicht z. B. <quote> und <Zitat> im selben Text zu haben.

Komponenten des Textes sind nach diesem Verständnis die Code Points, die Strukturinformation, die Anzeigeeinformation, die Regeln für das Markup, Metadaten und ein *Parser*, der nach dem EVA-Prinzip⁵⁷ den Text ‚errechnet‘ und zur Anzeige bringt bzw. zu einem Resultat verarbeitet. Wir haben also Daten, Datenformate, Anweisungen für Struktur und Layout (Programme, Skripte) sowie einen Prozessor, die insgesamt den Text konstituieren.⁵⁸ Diese kleinteilige Sicht ist nötig, um den spezifischen Textbegriff zu fixieren, der die Grundlage für eine digitale Text- und so auch Literaturwissenschaft bildet. ‚Text‘ in diesem Sinne ist nichts Statisches, Unidirektionales, das von einem Autor ins Werk gesetzt und von einem Leser gelesen wird, sondern prinzipiell etwas Multifunktionales und Prozedurales, das auch die Maschine als Leser kennt, dessen Autor durch die Anwendung von eigenen Skripten auch Leser ist⁵⁹ oder das in kombinatorischem Sinne als *Enhanced Publication* mehrdimensionale Zugänge eröffnet.

⁵⁵Vgl. Pédauque (Anm. 33).

⁵⁶World Wide Web Consortium, *Extensible Markup Language (XML)* (8. Dezember 1997), <https://www.w3.org/TR/PR-xml-971208.html> (letzter Aufruf 21.6.2018), hier: Abschn. 2.4, *Character Data and Markup*.

⁵⁷Eingabe – Verarbeitung – Ausgabe.

⁵⁸Thomas Stäcker, „Von Alexandria lernen‘ – die Forschungsbibliothek als Ort digitaler Philologie“, in: *Frauen – Bücher – Höfe. Wissen und Sammeln vor 1800. Essays in honor of Jill Bepler*, Wiesbaden 2018, 93–104.

⁵⁹Man denke an McCartys Bonmot vom „end maker“ statt „end user“ (Willard McCarty, *Humanities computing*, Basingstoke, Hampshire [u. a.] 2005, 15).

Aus diesem Textmodell ergeben sich für Bibliotheken und die Literaturwissenschaft Konsequenzen. Zunächst ist festzuhalten, dass im Falle der (Retro-) Digitalisierung die Charakteristika des Druckzeitalters in digitalen Code überführt werden müssen. Diesen Prozess kann man je nach Tiefe und Ausrichtung *Datafication* oder auch *digitales Edieren* nennen – in der Schrittfolge Image-digitalisierung, manuelle oder automatische Transkription, strukturelle und semantische Annotation sowie Interpretation bzw. Kommentierung. Im Falle einer *born-digital*-Publikation müsste ein Autor Anforderungen an ein digitales Dokument unter den gewandelten Bedingungen beachten, z. B. in der Frage, ob Strukturelemente eindeutig interpretierbar sind, ob Standards eingehalten werden oder ob Metadaten verfügbar sind. In einigen Bereichen haben sich bereits meist XML-basierte Formate als Standards durchgesetzt, die diese Anforderungen weitgehend erfüllen, wie JATS, JTEI oder DITA. XML ist zwar nur eine von mehreren möglichen Verfahren zur Textkodierung – z. B. spricht prinzipiell nichts dagegen, Text als Graph oder in JSON darzustellen –, doch scheint es trotz einiger Mängel⁶⁰ so weit durchgesetzt, dass eine grundsätzliche Neuorientierung wenig sinnvoll erscheint und man eher darauf sehen sollte, inwieweit sich die verschiedenen Konzepte komplementieren, insbesondere mit Blick auf graphentheoretische Ansätze, zu denen auch RDF⁶¹ gehört. Ohne auf diese Elemente des digitalen Textes umfassend eingehen zu können, ist doch die funktionale Unterscheidung von ‚Text als Sequenz von *Character Data*‘ und ‚Text als Kombination von *Character Data* und Markup‘ festzuhalten, die alle modernen Konzepte von digitaler bzw. digitalisierter Literatur durchzieht und die in ein Verhältnis von Funktion (Algorithmus) und Argument (Daten+Struktur) eingebettet ist. Aristotelisch kann man sie sich auch als *dynamis* und *energeia* vorstellen. Daten sind als Relationsbegriff in diesem Zusammenhang entweder *Code Points* als Gegenstand eines Textrenderingprozesses oder *Character Data* als Gegenstand eines Markierungsprozesses, der die Bedeutung des Textes maschinenlesbar macht.

Diese extrem funktional-technische und kleinteilige Interpretation von ‚Text‘ führt unmittelbar zu einem Gefühl des Ungenügens, vor allem mit Blick auf den sinnstiftenden Leseakt und die Bedeutung des Textes, doch ist sie nötig, um die für die digitale Sphäre relevanten Schichten unterscheiden und auch Bedeutung und Sinn kodieren zu können. Für diesen Schritt der Transformation hat sich insbesondere der Begriff ‚Modellierung‘ als nützlich erwiesen.⁶² Pedauque hat, ohne im vollen Umfang die Konsequenzen daraus zu ziehen, darauf hingewiesen, dass das Verfassen von Texten einen Vertrag zwischen Autor und seinem Leser

⁶⁰ Auf die Kritik am OHCO-Modell von XML und dem Problem des *Overlapping Markups* kann hier nicht eingegangen werden.

⁶¹ <https://www.w3.org/RDF/> (letzter Aufruf 21.6.2018).

⁶² Siehe insb. McCarty (Anm. 59).

impliziert.⁶³ Das kann bei digitalen Dokumenten bedeuten, dass der Autor ein genaues Verständnis auch der technischen Dimension des digitalen Textes haben muss, um seinen maschinellen und natürlichen ‚Leser‘ zu erreichen – oder aber, dass die Autorintention hinter den eigentlichen Modellierungsprozess zurücktritt und dem Leser eine weit prominentere Rolle in der ‚Interpretation‘ und Auswertung des Textes zukommt.⁶⁴ Autor- und Leserfunktion ändern sich, denn wenn ein Text geschrieben oder umgewandelt wird, um maschinenlesbar zu sein, müssen Elemente in den Text eingebracht oder mit ihm verbunden werden, die nicht für einen menschlichen Leser, sondern vor allem für eine Maschine ‚lesbar‘ sind. ‚Lesbarkeit‘ bedeutet in diesem Sinne nicht nur die Fähigkeit, einen Text auf der Basis von Sprache und Schrift zu verstehen, sondern auch ‚technische Lesbarkeit‘ auf der Basis von Daten und Struktur – mit anderen Worten: Digitale Texte müssen Schnittstellenanforderungen erfüllen (s. u.). Auch wenn die mathematische und statische Textauswertung kein neues Phänomen in der Literaturwissenschaft ist,⁶⁵ ist es doch die direkte Umwandlung (Modellierung) der Dokumente in Texte, die zum Zwecke integraler oder externer neuer Funktionen angereichert (annotiert) werden. Mit dem semantischen Markup erhält die Literaturwissenschaft zudem ein editorisches bzw. philologisches Werkzeug an die Hand, das es ihr ermöglicht, im Zuge einer interpretierenden Modellierung des Textinhalts genauer und expliziter zu sein, als es im analogen Feld möglich war.⁶⁶

Aus der Darstellung sollte deutlich werden, dass die Textwissenschaft bzw. dass das *Textual Scholarship* zu einer digitalen Textwissenschaft erweitert werden muss, die sich u. a. als digitale Philologie oder Computerphilologie der textlichen Datenverarbeitung und -repräsentation sowie Inhaltsmodellierung annimmt. Vonseiten der Remediatisierung ist sie nichts anders als digitale Editorik, von Seiten der *born-digitals* digitales Publizieren bzw. maschinenkonformes Schreiben. Für die Bibliothek stellt sie eine neue Sicht auf Literatur dar, die bisher in der Bibliothekspraxis kaum Berücksichtigung gefunden hat, und bestimmt den Text als ein Kompositum von Code Points, Strukturinformation, Anzeigeinformation, Regeln für das Markup, Metadaten und einem *Parser*, der für das ‚Ins-Werk-Setzen‘ bzw. die Darstellung sorgt. Konkret handelt es sich um eine Reihe von Dateien, wie z. B. .xml, .xslt, .xsd, .css oder.html, die zusammen ein funktionales Ganzes

⁶³ Pédauque (Anm. 33).

⁶⁴ Auch in der traditionellen Literaturwissenschaft hat es diese dominante sinnstiftende Leserfunktion schon gegeben, vgl. Arnold/Sinemus (Anm. 13), 102 ff.

⁶⁵ Walther L. Fischer, „Mathematik und Literaturtheorie“, in: *Archimedes* 22 (1970), 19–26; siehe auch ders., „Mathematische Texttheorie“, in: Arnold/Sinemus (Anm. 13), 44–61.

⁶⁶ McGann (Anm. 52), 90: „As we lay foundations for translating our inherited archive of cultural materials, including vast corpora of paper-based materials, into digital depositories and forms, we are called to a clarity of thought about textuality that most people, even most scholars, rarely undertake.“ „Translating“ heißt hier nichts anderes als ‚Edieren‘. Vgl. in diesem Zusammenhang auch seinen Begriff des Textes als „autopoietic event“ (108).

ergeben und in diesem Ensemble nicht nur die Stelle des Dokumentes einnehmen, sondern auch die stabile Grundlage jeder Langzeitarchivierung bilden.

3 Die Bibliothek: Sammeln – Erschließen – Benutzen

Die Aufgaben der Bibliothek werden funktional von alters her durch die Trias ‚Sammeln, Erschließen und Benutzen/Vermitteln von Literatur‘⁶⁷ beschrieben. Diese hat bis heute Bestand und kann als Leitlinie dienen, um die sich wandelnde Beziehung von Bibliothek und Literaturwissenschaft zu beleuchten. Doch müssen diese Begriffe mit der digitalen Wende neu interpretiert, die Bibliothek an die Anforderungen, die sich aus dem neuen Textbegriff ergeben, angepasst werden.

3.1 Sammeln

Bis vor wenigen Jahren war unstrittig, dass Bibliotheken Literatur in Form von Büchern oder ‚Medieneinheiten‘ sammeln, um der Literaturwissenschaft das Reservoir zu schaffen, das sie für ihre Forschungen benötigt.⁶⁸ Mit der digitalen Wende stellen sich aber die Fragen nach der Art und Form der Sammlung und nach dem Sammlungsort neu. Denn einerseits haben sich mit den Möglichkeiten des Internets die klassischen Publikationswege geändert, andererseits liegt auf der Hand, dass physische und *digitale Bibliothek* nicht mehr zusammenfallen müssen, auch wenn es aus organisatorischen Gründen pragmatische Gründe dafür geben mag. Eine *digitale Bibliothek* ist ihrer Natur nach ortlos, was schon daran abzulesen ist, dass man ohne Weiteres alle relevanten literaturwissenschaftlichen Werke zumindest im Textformat auf eine Festplatte bringen könnte, gewissermaßen als *library to go*. Das heißt freilich nicht, dass sich die Bibliothek als solche erübrigte, nur braucht man zum Sammeln keine Magazine und damit keinen konkreten Ort im üblichen Sinne mehr.⁶⁹ Das hat im Umkehrschluss dazu beigetragen, dass Bibliotheken es trotz vereinzelter mahnender Stimmen⁷⁰ versäumt haben, ihren Sammelauftrag auch auf nicht-trägerbasierte elektronische Medien auszudehnen und sie es bei der Schaffung von Zugang

⁶⁷ ‚Literatur‘ meint hier im Allgemeinen all das, was publiziert wurde.

⁶⁸ Zur Bedeutung des Sammelns in der Literaturwissenschaft siehe Wegmann (Anm. 6), 177 f. mit Literaturhinweisen.

⁶⁹ Natürlich hat auch das Digitale eine Ortseigenschaft, die gewöhnlich durch einen *Uniform Resource Locator* angezeigt wird. Da das Web aber prinzipiell Transklusion ermöglicht, ist nicht gesagt, dass auch alle Dateien bzw. Komponenten des Dokumentes an diesem Ort sein müssen. Vgl. auch Baum (Anm. 46).

⁷⁰ Doina Oehlmann, „Lizenzen oder Texte, Nutzung oder Hosting? Können Bibliotheken ihren Auftrag in Zeiten elektronischer Texte weiterhin ausreichend erfüllen?“, in: *ZfBB* 59 (2012) 5, 231–235.

haben bewenden lassen. Die Gründe lagen vor allem in den fehlenden Geschäftsgängen, mangelndem technischem Know-how, in bequem zugänglichen Online-Angeboten von Verlagen, die eine Eigenarchivierung überflüssig erscheinen ließen, und letztlich auch den schwierigen rechtlichen Rahmenbedingungen, nach denen nicht-trägerbasierte Medien nicht mehr als Sache behandelt werden konnten und lizenziert werden mussten.⁷¹ Insgesamt scheint eine Art Grundvertrauen zu herrschen, dass Nationalbibliotheken oder andere zentrale Einrichtungen diese Funktion übernehmen werden. Ob tatsächlich die gesamte Bandbreite der international erscheinenden e-only-Publikationen ‚irgendwo‘ gesammelt wurde und wird, ist aber trotz der Einrichtung von DFG-Sondersammelgebieten bzw. Fachinformationsdiensten zumindest offen. Dahinter steht auch die Frage, wer dieserart Archivaufgaben im digitalen Sektor überhaupt wahrnimmt und warum eine Bibliothek es tun sollte, zumal es nicht nur um die lizenzbewehrten, sondern auch freien Publikationen im *Open Access* geht, wo das Sammeln augenscheinlich nicht mehr dem unmittelbaren Zweck der Benutzung vor Ort dient. Nur lizenzierte Literatur benötigt geschützte Räume und durch Schrankenbestimmungen des Gesetzgebers abgesicherte Vermittlungsinstanzen wie Bibliotheken. Literatur, die im *Open Access* zugänglich ist, kann, so hat es den Anschein, in Zeiten von *Google* und Co. leicht von Zuhause aus benutzt werden – und es bedarf keiner sammelnden Einrichtung mehr. So sehr diese Evidenz auch einleuchten mag, erweist sie sich bei genauerem Hinsehen als trügerisch. Denn die Beschaffung und Nutzung von Texten im Netz stellt in der Praxis den Forscher vor große, teils unüberwindliche Herausforderungen und ist wenig geeignet, wissenschaftsfreundliche digitale Nutzungsszenarien zu entwickeln. Texte im Netz sind oft von unbekannter Quelle, nicht zuverlässig erreichbar und zitierbar und, da dekontextualisiert, schwer einzuordnen. In vielen Fällen ist weder deutlich, auf welcher Textgrundlage sie beruhen, noch weiß man um ihre Qualität oder Vollständigkeit. Auch das Vorhandensein ausreichender Metadaten steht infrage – Mängel, die für eine digital arbeitende Literaturwissenschaft beträchtliche Hürden darstellen. Aus wissenschaftlicher Sicht ist es weniger wichtig, dass Texte irgendwo im Netz heruntergeladen werden können, als dass sie als differenziert beschriebene, integrale Teile einer wohldefinierten Sammlung verstehbar werden und eingeordnet werden können.

Unter diesem Gesichtspunkt gewinnt der Sammlungsakt sogar an Bedeutung, da größere Anstrengungen erforderlich sind, ein einheitliches und konsistentes Korpus bzw. eine digitale Sammlung⁷² herzustellen – eine Aufgabe, die neu für Bibliotheken ist, da nicht nur weitgehend einheitlich gestaltete Bücher zum Wiederfinden mit Signaturen versehen und magaziniert, sondern zunächst die

⁷¹ Erst mit der jüngsten Urheberrechtsreform, die zum 1. März 2018 in Kraft trat, ist die Archivierungsfrage wieder eindeutig geregelt.

⁷² Das Sammlungsthema wird deswegen nicht von ungefähr im *DH*-Kontext prominent diskutiert, siehe Christof Schöch, „Aufbau von Datensammlungen“, in: Fotis Jannidis/Hubertus Kohle/Malte Rehbein, *Digital Humanities. Eine Einführung*, Stuttgart 2017, 223–233.

‚Bücher‘ (Dokumente *qua* Dateien) in eine sammlungsgeeignete, d. h. einheitliche Form, überführt werden müssen. Das Sammeln selbst verändert sich auf diese Weise in mehrfacher Hinsicht, zumal wenn es – wie perspektivisch zu erwarten – ausschließlich unter der Maßgabe von *Open Access* stattfindet. Digitale Sammlungsbildung kann z. B. bedeuten, dass eine Sammlung sich bei der anderen in einem freien Copy-&-Paste-Prozess ‚bedient‘, nicht unähnlich der ‚Kopiermaschine‘ Alexandrias.⁷³ Es kann bedeuten, dass die Kriterien, die bislang sammlungskonstitutiv waren, etwa die Figur der Sammlerpersönlichkeit, sich relativieren und Begriffe wie ‚seine Bibliothek‘ oder die ‚Sammlung der Bibliothek XY‘ eine andere oder gewandelte Bedeutung erhalten. Es kann auch bedeuten, dass Dokumente sammlungsbezogen modifiziert bzw. angereicht und nicht wie bisher *as-is* archiviert werden. An organisatorischen Änderungen tritt hinzu, dass die, was Texte (*Character Data*+Struktur) anlangt, schier unendlichen Speichermöglichkeiten eine Erwerbungs Auswahl unter dem Gesichtspunkt des Magazinplatzes ebenso obsolet erscheinen lassen wie die Makulierung redundanter oder überflüssig scheinender Literatur, die Platz schafft für Neuerwerbungen. Auch die Erwerbungskosten konvergieren gegen Null. E-Books und E-Journals im *Open Access* kosten zunächst nichts, sie müssen nicht gekauft, sondern nur gefunden und heruntergeladen werden.

Doch täusche man sich nicht. Es entstehen für Bibliotheken, die sich dieser Aufgabe annehmen, beträchtliche neue Kosten im Verwaltungsprozess und auch in der Langzeitarchivierung und -verfügbarkeit,⁷⁴ die eine der wesentlichen Aufgaben bibliothekarischen Sammelns bleiben. Allerdings hat es Sinn, sie an zentralen regionalen oder nationalen Stellen zu konzentrieren und sie von der jeweiligen ‚Bibliothek‘ lediglich administrieren zu lassen. Erste Modelle sind in Entwicklung.⁷⁵ Sammeln unter digitalen Bedingungen reicht aber noch weiter. Wenn die Publikation im Internet im *Open Access* unmittelbar keine vermarktende Instanz mehr benötigt, weil theoretisch jeder Wissenschaftler selbst seine Beiträge publizieren kann, und wenn es umgekehrt Aufgabe der Bibliothek ist, wissenschaftliche Publikationen zu sammeln, anzubieten und zu vermitteln, dann liegt der Schluss nahe, dass Bibliothek und Wissenschaft in diesem Segment Hand in Hand arbeiten sollten, um das Publikationsinteresse der einen und das Sammlungs- und Vermittlungsinteresse der anderen Seite direkt miteinander zu verzahnen und aufeinander abzustimmen. Die Bibliothek weitet damit zugleich automatisch ihren Nutzerkreis aus, selbst wenn ihre Zielgruppe die eigene Klientel bleibt. Sammeln ist daher vonseiten der Bibliothek zugleich eine Unterstützungsleistung bei der Herstellung und Vermittlung wissenschaftlicher Publikationen

⁷³ Stäcker (Anm. 58).

⁷⁴ Zu diesem Thema siehe insb. Nestor (www.langzeitarchivierung.de [letzter Aufruf 21.6.2018]) und das *Open Archival Information System (OAIS)* bzw. ISO 14721:2012.

⁷⁵ Vgl. TIB Hannover, SLUB Dresden oder die zentrale Archivierung in NRW.

unter nicht-kommerziellen Aspekten,⁷⁶ mit anderen Worten: Die Bibliothek wird zum wissenschaftlichen Publikationsdienstleister und Unterstützer eines rein an wissenschaftlichen Kriterien orientierten und nicht durch kommerzielle Interessen unterbrochenen oder gefährdeten *Research Data Cycle* bzw. *Text Life Cycle*.⁷⁷

Diese allgemeinen Charakteristika betreffen zwar nicht nur die Literaturwissenschaft, haben aber für sie beträchtliche Auswirkungen. Wenn ‚Sammeln‘ immer weniger nur eine Sache des ‚Ablegens‘ und Bereitstellens, sondern auch des ‚Herstellens‘ ist, übernimmt die Bibliothek für die *digitale Literaturwissenschaft* eine produktive Aufgabe nicht nur in der Bereitstellung von Texten und ihrem ‚Erwerb‘, sondern auch darin, dass sie den Rahmen bietet zu deren Herstellung – in dem oben genannten zweifachen Sinn als (1.) *Datafication* der Transformation analoger Literatur in digitale im Editionsprozess und (2.) der Erzeugung von *born-digitals*, also von Texten der literaturwissenschaftlichen Fachliteratur, die sich mit den Gegenständen dieser Literatur in der Bibliothek treffen. Bibliotheken sind daher als Orte ‚digitaler Sammlungen‘ auch virtuelle Forschungsräume, deren sich die Literaturwissenschaft bedienen kann, um ihre Ergebnisse zu publizieren. Die Bibliothek als wissenschaftsimmanente Forschungsinfrastruktur gedacht,⁷⁸ muss jedoch Sorge tragen, dass die wissenschaftsimmanenten Qualitätssicherungsverfahren der Literaturwissenschaft gewährleistet sind und Strukturen entstehen, durch die die Bibliothek als wissenschaftlicher Publikationsdienstleister der Literaturwissenschaft attraktive Angebote machen kann, die auch Akzeptanz finden und wissenschaftliches Prestige genießen.⁷⁹ Sammlungsbildung bedarf in dieser Hinsicht auf mehreren Ebenen einer Neuausrichtung und einer strengeren Orientierung an qualitativen und auch differenzierenden Angeboten. Zugleich ist eine Neugestaltung des althergebrachten Fachreferats notwendig, dessen Aufgabe es sein muss, Sammelentscheidungen einerseits nach philologisch-bibliographischen Kriterien zu treffen und digitale Texte zu übernehmen oder herzustellen, die den fachlichen Anforderungen und digitalen Rahmenbedingungen entsprechen (qualitätsgesichert bzw. *peer-reviewed*, bibliographisch identifizierbar, unter offenen Lizenzen, mit Beigabe oder verlässlicher Verlinkung von Quellen bzw. Forschungsdaten etc.), andererseits nach technisch-formalen Kriterien, wo zu entscheiden ist, ob erworbene Ressourcen für digitale Anwendungen genügen oder aufbereitet werden müssen (Unicode, definierte Medientypen, bibliographische und strukturelle Metadaten inkl. Linkstrukturen etc.). Aufbereitete offene

⁷⁶Die Produktion von Literatur – auch wissenschaftlicher –, die mit dem Ziel der Gewinnerzielung geschrieben wird, fällt nicht in den Aufgabenbereich einer akademischen Einrichtung.

⁷⁷Johanna Puhl/Peter Andorfer/Mareike Höckendorff u. a., *Diskussion und Definition eines Research Data LifeCycle für die digitalen Geisteswissenschaften. DARIAH-DE Working Papers 11*, Göttingen 2015, <http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2015-4-4>.

⁷⁸Kathleen Fitzpatrick, *Planned obsolescence. Publishing, technology, and the future of the academy*, New York u. a. 2011.

⁷⁹Martin P. Eve, *Open access and the humanities. Contexts, controversies and the future*, Cambridge, UK, 2014, <https://doi.org/10.1017/CBO9781316161012>.

Sammlungen dieser Art, die eine signifikante Größenordnung erreicht haben, gibt es bis heute nur wenige,⁸⁰ und sie bilden im Bibliothekskontext immer noch die Ausnahme, wenngleich sie für die *digitale Literaturwissenschaft* wichtige Ressourcen wären.

3.2 Erschließen

Größere Korpora erlauben oft keine Differenzierung nach Textsorten und machen eine trennscharfe digitale Analyse schwierig. Zugleich ist es angesichts der mannigfachen Textvarianten und -formen im Netz eine besondere Herausforderung, daraus konsistente Sammlungen zu kreieren. Ein Beispiel dieser Probleme liefert *TextGrid*, das den lobenswerten Versuch unternommen hat, Texte aus heterogenen Quellen nach Gattungen zu klassifizieren, strukturell zu vereinheitlichen und zur Nachnutzung in Editionen oder zum *Textmining* zur Verfügung zu stellen. Die aus Zeno.org und anderen Quellen übernommenen Texte und Strukturdaten wurden zu diesem Zweck nach TEI-XML überführt und, so weit wie möglich, vereinheitlicht.⁸¹ Allerdings waren die benötigten Ressourcen beträchtlich, sodass das Korpus bis heute nicht vollständig bereinigt ist. Der Fall zeigt, dass die Remedialisierung bzw. analog-digitale Konversion größerer Korpora die Leistungsfähigkeit einzelner Einrichtungen übersteigt und die meist projekthafte Organisation ungeeignet ist, zu einem wirklich flächendeckenden Angebot zu kommen. Eine nationale Erschließungsagenda, die das Ziel der einheitlichen Bereitstellung von Texten verfolgt, ist daher unverzichtbar. Erschließung bewegt sich hier nicht mehr nur in der traditionellen bibliothekarischen Unterscheidung von Formal- und Sacherschließung, sondern schließt die Aufbereitung der Quellen selbst in der Remedialisierung und die Beachtung von Standards in der Herstellung von Publikationen ein. Insofern werden die deskriptiven Prozesse von der Metadatenebene auf die Daten-/Dokumentenebene bzw. Textebene erweitert. Analoge Materialien werden für eine digitale Nutzung dadurch ‚erschlossen‘, dass man einen Konversionsprozess implementiert, der von der Imagedigitalisierung über die Volltexterfassung und -konversion⁸² bis zur Strukturdatenauszeichnung und Edition führt. Wichtige nationale Agenden in dieser Hinsicht sind die DFG-Projekte zur Digitalisierung der Verzeichnisse der im deutschen Sprachraum erschienenen Drucke des 16., 17. und 18. Jahrhunderts

⁸⁰Eine Ausnahme bildet z. B. das Deutsche Textarchiv: <http://www.deutschestextarchiv.de/> (letzter Aufruf 21.6.2018).

⁸¹<https://textgrid.de/digitale-bibliothek> (letzter Aufruf 21.6.2018).

⁸²Thomas Stäcker, „Konversion des kulturellen Erbes für die Forschung. Volltextbeschaffung und -bereitstellung als Aufgabe der Bibliotheken“, in: *o-bib* 1 (2014), 220–237, <https://doi.org/10.5282/o-bib/2014H1S220-237>.

(VD 16, VD 17 und VD 18)⁸³ und das unlängst begonnene Vorhaben zur *OCR-D*.⁸⁴ Darauf aufbauend, stehen Maßnahmen der systematischen Vereinheitlichung (Strukturdaten) der Texte, für literaturwissenschaftliche Quellen auf der Basis der TEI, zunächst im Sinne einer Rohkodierung, die bis zu historisch-kritischen Editionen ausgebaut und angereicht werden können.⁸⁵

Der digitalisierte oder digitale Text bildet den Unterbau für die klassische Formal- und Sacherschließung. An dieser ändert sich einerseits wenig – auch digitale Dokumente benötigen Metadaten zu Autor, Titel, Erscheinungsort und -jahr, und ihr Inhalt muss klassifiziert und beschrieben werden –, andererseits bringt die digitale Form Änderungen bei der Erschließung mit sich. Die Formal- und Sacherschließung dient nach dem aktuellen Katalogisierungsstandard RDA dem Finden („d. h. Ressourcen finden, die den vom Benutzer festgelegten Suchkriterien entsprechen“), dem Identifizieren („d. h. bestätigen, dass die beschriebene Ressource der gesuchten entspricht, oder zwischen mehreren Ressourcen mit ähnlichen Eigenschaften unterscheiden“), dem Auswählen („d. h. eine Ressource auswählen, die den Bedürfnissen des Anwenders entspricht“) und dem Zugang erhalten („d. h. die beschriebene Ressource erwerben oder Zugang zu ihr erhalten“).⁸⁶ Da die Bibliothek sich im Digitalen entgrenzt, besteht eine besondere Herausforderung darin, die bibliothekarische Erschließung literaturwissenschaftlicher Quellen und Forschungsliteratur in einem Webkontext zu denken, d. h. beispielsweise Eindeutigkeit möglichst auch auf einem internetweiten Level sicherzustellen und Zugänge unter Berücksichtigung multikultureller Aspekte und sprachlicher Vielfalt zu gestalten. Vor diesem Hintergrund ändert sich die Form dieser Erschließung und verlangt nach anderen webbasierten Techniken. Dazu gehört nicht nur die differenzierte Beigabe von und Anreicherung mit Metadaten, die sich in deskriptive, strukturelle, administrative und technische unterscheiden lassen, sondern auch semantische Beschreibungssprachen wie RDF, RDFS oder OWL, die Inhalte in maschinenlesbarer Form charakterisieren und Beziehungen der Texte und Dokumente untereinander mit Hilfe von Graphen explizit machen. Gerade die Perspektive eines *Semantic Web* könnte in einer Utopie aller semantisch verbundenen Ressourcen des Netzes dazu führen, dass strukturierte Texte und Daten, wie in der sogenannten *Semantic Web Layer Cake*⁸⁷ dargestellt, wirklich unabhängig von proprietären Anwendungen und

⁸³ Digitalisate wurden mit den Datenbanken verknüpft: www.vd16.de, www.vd17.de und www.vd18.de (letzter Aufruf der Links 21.6.2018).

⁸⁴ www.ocr-d.de (letzter Aufruf 21.6.2018).

⁸⁵ Vgl. auch die DFG-Förderkriterien zu literaturwissenschaftlichen Editionen: http://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/foerderkriterien_editionen_literaturwissenschaft.pdf (letzter Aufruf 21.6.2018).

⁸⁶ *Ressource Description Access*, <http://access.rdatoolkit.org/> (Lizenz erforderlich), hier: „Einführung“, „0.0 Ziel und Geltungsbereich“.

⁸⁷ *Wikipedia*, „Semantic Web Stack“, https://en.wikipedia.org/wiki/Semantic_Web_Stack (letzter Aufruf 21.6.2018).

technischen Besonderheiten für alle Arten von Nutzungen zur Verfügung stehen und intertextuelle Forschungen auf der Ebene des Internet vorangetrieben werden. Allerdings steckt die Anwendung von *Semantic-Web*-Techniken in den Literaturwissenschaften noch in den Kinderschuhen, und eine kooperative bibliothekarisch-literaturwissenschaftliche Modellierung von Textbeziehungen bzw. von deren semantischen Komponenten von Texten bedarf noch einer breiteren Erprobung.⁸⁸ All diese Arbeiten gehen über die traditionellen Katalogisierungsarbeiten hinaus, gehören aber in den Komplex der Erschließung digitaler literaturwissenschaftlicher Ressourcen.

Es ist deutlich, dass auch das Erschließen der Bibliothek einen neuen Charakter erhält. Es ist nicht mehr länger nur rezeptiv, nimmt fertige Bücher bzw. Dokumente entgegen und katalogisiert sie, sondern bereitet sie auch in vielfältiger Weise auf, reichert sie an und vernetzt sie. Das Ziel der Erschließung digitaler Medien liegt darin, eine möglichst homogene, konsistent strukturierte Text- und Datenmenge für die Recherche, Visualisierung und digitale Nachnutzung vor allem im Internet zur Verfügung zu stellen. Dies geschieht nicht mehr nur auf der deskriptiven Ebene der Sammlungen und Dokumente, sondern auch auf jener der Texte selbst (z. B. durch Auszeichnung nach TEI, *POS Tagging*, NEP etc.). Erschließung muss dabei nicht unbedingt bedeuten, dass diese manuell erfolgt, man kann auch an automatisierte Verfahren denken. Beispielsweise ließe sich *Topic Modeling* auch zur Klassifikation von Dokumenten einsetzen. Der Text selbst wird zu einer Art Werk- oder Rohstoff, den man je nach Nutzungsszenario auch mit dem Begriff ‚Forschungsdaten‘ belegen kann. Durch Normalisierung und Standardisierung von Texten, die zusätzlich zur Originalform – also gleichsam als Derivat – erzeugt werden, ist sichergestellt, dass die Sammlungsbildung als die Grundlage für alle literaturwissenschaftliche Forschungsfragen erfolgt und Daten über definierte Schnittstellen (s. u.) ausgeliefert werden können. Insofern wird es zu einer zentralen bibliothekarischen Erschließungsaufgabe, Texte und Daten, die bisher in proprietären Formaten und in abgeschlossenen, von außen unzugänglichen Datenbanken (Datensilos) vorliegen, zu öffnen, um sie so auch im Internet zugänglich zu machen.

3.3 Benutzen

Die reine Zurverfügungstellung von Texten ist angesichts der fortschreitenden Digitalisierung und freien Bereitstellung auch von Forschungsliteratur von abnehmender Relevanz. Während der Zugang zu Literatur im Mittelalter und noch in der Frühen Neuzeit auch unabhängig von der Lesefähigkeit ein Privileg Weniger war, die Zugang zu Büchern oder Sammlungen erhielten, ist die Verfügbarkeit heute zu fast allen literarischen Werken der Literatur kaum noch

⁸⁸Vgl. die Bereitstellung von Daten zu Emblembüchern der HAB als LOD: <https://old.datahub.io/dataset/emblem> (letzter Aufruf 21.6.2018).

beschränkt und in vielen Fällen auch vom heimischen Schreibtisch aus möglich. Klassische Bibliotheksaufgaben der Benutzung wie Aus- und Fernleihe, Bereitstellung von Lesesaalflächen oder physische Buchaufstellung werden daher für die Forschungsliteratur im engeren Sinne an Bedeutung verlieren und allein für die Konsultation von gedruckten Originalen und Sondersammlungen relevant bleiben.⁸⁹

Eine wesentliche Funktion der Bibliothek liegt im Suchen und Finden, die Sammeln und Erschließen zu ihrer Voraussetzung haben. Zugänge zu Werken der Literatur haben sich diversifiziert und finden nicht mehr exklusiv über Bibliothekskataloge oder Fachbibliographien statt – auch wenn sie in den meisten Fällen die Datengrundlage dafür bieten –, sondern über zentrale, leicht bedienbare Suchmaschinen, insbesondere *Google*. Im Milieu des digitalen Überflusses, aber auch der Unübersichtlichkeit, wächst die Bedeutung differenzierter Zugangswege zu Texten und der Wunsch der Forschung nach qualitativ hochwertigen und konsistenten Sammlungen, die es in einem Webkontext erlauben, schnell zu ‚relevanten‘ Ressourcen zu kommen. *Google* und andere Suchmaschinen leisten zwar Beträchtliches, doch bleiben spezialisierte Zugänge auch weiterhin unverzichtbar, sowohl aus Gründen der bibliographischen Identifizierbarkeit und gezielter Auswahl als auch aus wissenschaftsmethodischen Gründen, denn die Funktion der Suchalgorithmen bleibt eine Unbekannte und damit eine Bewertung der Heuristik von Suchmaschinen weitgehend unmöglich. Im Zweifelsfall wirken Suchalgorithmen, die für eine wissenschaftliche Nutzung ungeeignet sind.⁹⁰ Vor diesem Hintergrund verschiebt sich auch die Funktion der Bibliothek vom *Gatekeeper* hin zum Filter – ein Konzept, dem auch moderne Modelle des *Post-Publication Reviewing* folgen.⁹¹ Sie macht durch gezieltes Herausfiltern die ungeordnete Informationsmasse allererst zugänglich und ist Teil eines Rezeptionsprozesses, der eine Publikation als forschungsrelevant herausstellt. Wie früher nur die wichtigen Bücher ‚gekauft‘ wurden, werden jetzt nur die wichtigen Bücher aus einem ubiquitären freien Datenmeer ‚gefiltert‘.

Im digitalen Kontext erweitern sich zudem die Anforderungen an die Such- und Nutzungsmöglichkeiten, da es nicht nur darum geht, Titel zu finden, sondern auch Texte zu analysieren. Typische Textmining-Verfahren (*Topic-Modelling*, *Stilometrie*, *Sentiment Analysis* etc.) erfordern nicht nur die Identifikation von Ressourcen, sondern auch Text- und Datenhoheit sowie Transparenz bei den Daten- und Korpusstrukturen, sodass die Benutzung immer auch auf digitaler Sammlungsbildung und gegenstandsadäquater Erschließung aufbaut. Ohne

⁸⁹D. h. nicht, dass die Bibliothek mit anderen Funktionen als physischer Ort relevant bliebe.

⁹⁰Zum Problem der Personalisierungen von Webseiten siehe Eli Pariser, *The filter bubble. What the Internet is hiding from you*, London 2011; vgl. auch die Kritik von Wegmann (Anm. 6), 314–315, an Software, die dem Leser immer nur seine eigenen Vorlieben zuspielt.

⁹¹Kathleen Fitzpatrick, „Peer Review“, in: Susan Schreibman/Ray Siemens/John Unsworth (Hg.), *A New Companion to Digital Humanities*, Chichester, West Sussex/Malden, MA/Oxford 2016, 439–448.

speziell an die Fragestellung angepasste Werkzeuge sind zudem wissenschaftliche Aussagen in vielen Fällen kaum möglich oder in ihrer Validität zumindest fragwürdig. Hier ergeben sich für die bibliothekarische Benutzung wichtige Arbeitsfelder, indem sie die *digitale Literaturwissenschaft* in der Anwendung der jeweiligen Werkzeuge berät, Unterstützung bei der Text- und Datenerzeugung leistet oder den Forschungsdaten- bzw. Textmanagementprozess begleitet. In der bibliothekarischen Benutzung geht es daher auch um Vermittlung digitaler und informationswissenschaftlicher Kompetenz zum Finden, Bearbeiten und Analysieren von Texten sowie die Unterstützung bei der vorbereitenden Erschließung und Publikation von Texten und Daten im Sinne eines literaturwissenschaftlichen Forschungsdatenmanagements. Wie früher vermittelt die bibliothekarische Benutzung den Zugang zur Literatur, zugleich wird sie aber durch die Spezialisierung in der Nutzung von digitalen Methoden und Werkzeugen in weiterem Umfang als bisher zum Partner der Forschung und übernimmt als Makler des Zugangs zu und der Modellierung von Texten eine wichtige Rolle im Forschungsprozess.

4 Die Bibliothek als Schnittstelle

Wegmann hat als einer der ersten Literaturwissenschaftler die Bibliothek unter dem Gesichtspunkt ihrer Schnittstellenfunktion thematisiert.⁹² Anders als die meisten älteren Medientheoretiker⁹³ nimmt er gegenüber einer ‚Schnittstelle‘ nicht eine hardwareorientierte Perspektive im Sinne eines Maschine-Maschine- oder Mensch-Maschine-Verhältnisses ein, sondern für ihn bezeichnet eine Schnittstelle „als umfassende technische Metapher die strukturelle Koppelung von Sinn-Kommunikation und apparativen Strukturen“. Und weiter: „Entscheidend ist, daß das Konzept der Schnittstelle eine interaktive Beziehung benennt“.⁹⁴ Technik und Bewusstsein stehen in einem sich wechselseitig determinierenden Verhältnis und eröffnen so neue ‚Schemata‘ des Suchens und Findens in der Bibliothek. Auch die neuere Medientheorie fasst den Schnittstellenbegriff weiter. Hookway hält fest: „The interface delimits a specific cultural space, within which a specific set of relations occur [...]. Its effects are registered not only in the opening up for access or experience of otherwise unavailable spatialities and temporalities, but also in how space and time are understood and treated within culture“.⁹⁵ Ähnlich wie Wegmann kommt er zu der Einsicht, dass das *Interface* den Nutzer nicht nur als

⁹²Wegmann (Anm. 6), Kap. 8: „Suchen und Finden. Über die Schnittstelle ins Labyrinth“, hier insb.: 306 ff.

⁹³Etwa in der Folge von Kittler, vgl. Sabine Wirth, „Between Interactivity, Control, and ‚Everydayness‘ – Towards a Theory of User Interfaces“, in: Florian Hadler/Joachim Haupt (Hg.), *Interface Critique*, Berlin 2016, 18 f.

⁹⁴Wegmann (Anm. 6), 307.

⁹⁵Branden Hookway, *Interface*, Cambridge, MA, u. a. 2014, 17.

Handelnden kennt, sondern diesen in seinen operativen Festlegungen immer auch beeinflusst und bestimmt. Die Schnittstelle ist für ihn nicht nur ein technisches Protokoll: „For while the interface might seem to be a form of technology, it is more properly a form of relating to technology“.⁹⁶ Die Schnittstelle tritt hier als medialer Übergangspunkt in Erscheinung, dem auch Eigenständigkeit und nicht nur transitorische Bedeutung als Vermittlungsprozess beigemessen wird. „The interface cannot be captured as a mere remediation of other media, but unfolds its own specific mediality“.⁹⁷ Auf dieser Abstraktionsstufe⁹⁸ werden Eigenschaften beschrieben, die man für die Rolle der Bibliothek operationalisieren kann, sodass sie nicht nur als eine Einrichtung bestimmbar ist, die im technischen Sinn Schnittstellen bereitstellt, sondern im Sinne dieser ‚Metapher‘ selbst als Schnittstelle zu den in ihr aufbewahrten (digitalen) Texten fungiert. Ihre Form und ihr Zuschnitt, ihre Struktur und Erschließung ist immer auf ihre Schnittstellenfunktion bezogen. Eine Bibliothek ist ein toter Bücherhaufen, wenn sie nicht mit ihren Benutzern kommuniziert und interagiert. So ist eine Schnittstelle zwar im technischen Sinne ein Protokoll zur Nachrichtenübermittlung – und die Bibliothek konkretisiert sich auch in diesen Protokollen –, aber sie bezeichnet zugleich auch eine allgemeine Übergangs- und Vermittlungsfunktion zwischen verschiedenen autonomen Seiten und Systemen. Und das nicht nur im Sinne von Nutzer und Bibliothek, sondern auch von Gestern und Heute, von Orient und Okzident,⁹⁹ von Geistes- und Naturwissenschaften usw. Die Schnittstellenfunktion ist als einer bestimmten Methodik folgender Prozess zu verstehen, den Wirth wie folgt beschreibt: „Building on the idea of the interface as a process, the term ‚interface‘ opens up the possibility to think the relation of human user and technology in its operability. The operability not only enables ‚interactivity‘ but first and foremost generates the possibility of it by providing a ‚ready-to-hand‘ space of possible activity“.¹⁰⁰ Dieser Raum ist letztlich als digitaler Raum zu verstehen und der Begriff der ‚Schnittstelle‘ ergibt nur in einem digitalen Kontext wirklich Sinn bzw. lässt sich nur dort angemessen umsetzen. Die gesammelten und erschlossenen digitalen Texte werden je nach Anforderung vielfältig modifiziert und in verschiedenen Formaten angeboten. Die diversen Extraktions-, Konversions-, Visualisierungs- oder auch sich in formalen – d. h. sich in maschinenlesbaren Sprachen ausdrückenden – literaturwissenschaftlichen Modellierungsprozesse, die Metadaten und Texte gleichermaßen umfassen,

⁹⁶ Ebd., 1.

⁹⁷ Manovich nach Sabine Wirth, „Between Interactivity, Control, and ‚Everydayness‘ – Towards a Theory of User Interfaces“, in: Florian Hadler/Joachim Haupt (Hg.), *Interface Critique*, Berlin 2016, 17–35, hier: 21.

⁹⁸ Kritik an der mangelnden Konkretion des Schnittstellenbegriffs bei Hookway und Galloway siehe ebd., 31: „sometimes appears to be interchangeable with the word medium“.

⁹⁹ Als Metapher des Raumes.

¹⁰⁰ Wirth 2016 (Anm. 93), 33.

lassen sich über die Schnittstellenfunktion der Bibliothek interpretieren. Die Bibliothek ist die Infrastruktur für Texte und Forschungsdaten, die das schriftliche kulturelle Erbe in eine maschinenlesbare, algorithmisch prozessierbare Form bringt, nachhaltig sichert und der *digitalen Literaturwissenschaft* zur Verfügung stellt, die ihrerseits Daten und Texte digital bearbeitet, um sie dann, vermittelt über Schnittstellen, zurückzuspielen. Wesentliche Komponenten dieses Prozesses sind Nachhaltigkeitskonzepte, Standardbildung, freie Zugänglichkeit durch *Open-Access*-Lizenzen und eine kooperative Wissenschaftskultur, die an die Stelle des traditionellen geisteswissenschaftlichen Einzelforschers Teams setzt, die in Forschungsprojekten die Informatik, Informations- und Bibliothekswissenschaft bzw. *Digital Humanities* sowie Fachwissenschaft in sich vereinigen.

Zusammenfassend lässt sich festhalten: Die Bibliothek wird als *digitale Bibliothek* zu einem virtuellen Forschungsraum, der sich im Begriff der Schnittstelle konkretisiert. Sie ist eine konkrete Funktion der Sammlung, Erschließung und Benutzung von Texten und Daten, nicht mehr ein Ort der Bücher. Sie sprengt frühere am Dokument haftende und dort fixierte Textformationen auf. Sie ist nicht nur ein Arbeitsplatz, an dem Bücher bereitgestellt werden, sondern wird zu einem integralen Element einer digital arbeitenden Literaturwissenschaft. Durch nach Qualitätskriterien bestimmtes Sammeln, Transformieren und Produzieren digitaler Texte und Daten sowie deren Langzeitsicherung bietet sie der *digitalen Literaturwissenschaft* zugleich den Gegenstand ihrer Forschung und ist im Digitalen ihr Bewegungs- und Beziehungsraum, der sich nicht in der Unbestimmtheit des Internets verliert, sondern einen zwar offenen, aber definierten digitalen Raum der Intertextualität bzw. Interaktivität bildet. Mithilfe des Textes als Code und Struktur, die als solche algorithmisch zu übersetzen, über technische Schnittstellen zugänglich und mit digitalen Methoden bearbeitbar sind, werden Anzeige- oder Lesevarianten erzeugt, verlässliche Suchen ermöglicht und Wege in unvordenkliche hypertextuelle Netze gebahnt. So verändert sich im Digitalen die Beziehung von Bibliothek und Literaturwissenschaft durch die gleichsam symbiotische Vereinnahmung der Literaturwissenschaft durch die Bibliothek – und umgekehrt der Bibliothek durch die Literaturwissenschaft. Die Bibliothek ist einerseits eine digitale Ressource, steht aber andererseits als ihr Element auch in einem gleichsam kybernetischen bzw. interaktiven Verhältnis zur Literaturwissenschaft, die hermeneutisch-zirkulär, mit digitalen Methoden arbeitend, ihre Forschungsergebnisse in den virtuellen Raum der Bibliothek der Texte einschreibt, als Hypertext, als Annotation, als Transkription und Edition oder als semantisches Netz. Zurückgebunden an die eingangs genannten Aspekte der Bibliothek lässt sich so die Beziehung der Bibliothek zur *digitalen Literaturwissenschaft* als prozesshaft gedachte Schnittstelle näher und neu bestimmen – als Laboratorium, Gedächtniseinrichtung, Ort der kulturellen Identität, humane Anstalt, Schatzhaus, Netzwerk oder auch Dienstleister.

Literatur¹⁰¹

- Arnold, Heinz L./Detering, Heinrich (Hg.), *Grundzüge der Literaturwissenschaft*, München 42001.
- Arnold, Heinz L./Sinemus, Volker (Hg.), *Grundzüge der Literatur- und Sprachwissenschaft*, Bd. 1: *Literaturwissenschaft*, München 61980.
- Assmann, Jan, *Das kulturelle Gedächtnis. Schrift, Erinnerung und politische Identität in frühen Hochkulturen*, München 1992.
- Barthes, Roland, „De l'œuvre au Texte“, in: *Revue d'esthétique* 3 (1971), 225–232.
- Baum, Constanze, „Im Hyperfluss“, in: Dies./Ulrike Gleixner/Jörn Münkner u. a. (Hg.), *Biographien des Buches*, Göttingen 2017, 326–348.
- Blume, Peter, „Bibliothekarische Systematiken und Fachsystematik(en) der Philologien – eine vielschichtige Beziehung“, in: Stefan Alkner/Achim Hölter (Hg.), *Literaturwissenschaft und Bibliotheken*, Göttingen 2015, 139–152.
- Bolter, Jay D./Grusin, Richard, *Remediation. Understanding new media*, Cambridge, MA, [u. a.] 2000.
- Brandner, Andreas, „Bibliotheken als Laboratorien der Literaturwissenschaft? Innenansichten analoger, digitaler und hybrider Wissensräume“, in: Stefan Alkner/Achim Hölter (Hg.), *Literaturwissenschaft und Bibliotheken*, Göttingen 2015, 115–138.
- Caton, Paul, „On the term ‚text‘ in digital Humanities“, in: *Literary and Linguistic Computing* 28 (2013), 209–220.
- Contzen, Adam, *Politicorum Libri Decem*, Köln/Mainz 1621.
- Coombs, James H./Renear, Allen H./DeRose, Steven J., „Markup Systems and the Future of Scholarly Text Processing“, in: *Communications of the ACM* 30 (1987), 933–947, DOI <https://doi.org/10.1145/32206.32209>.
- Derrida, Jacques, *Schrift und Differenz*, Frankfurt a. M. 1972 [frz. 1967].
- Deutsche Forschungsgemeinschaft, *Förderkriterien zu literaturwissenschaftlichen Editionen* 11 (2015), http://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/foerderkriterien_editionen_literaturwissenschaft.pdf.
- Eggert, Paul, „Text-encoding, Theories of the Text, and the ‚Work-Site‘“, in: *Literary and Linguistic Computing* 20/4 (2005), 425–435.
- Eichler, Ferdinand, *Begriff und Aufgabe der Bibliothekswissenschaft. Vortrag gehalten im Oesterreichischen Verein für Bibliothekswesen in Wien am 30. Mai 1896*, Leipzig 1896.
- Eichler, Ferdinand, *Bibliothekswissenschaft als Wertwissenschaft. Bibliothekspolitik als Welt-politik*, Graz/Leipzig 1923.
- Eve, Martin P., *Open access and the humanities. Contexts, controversies and the future*, Cambridge, UK, 2014, DOI <https://doi.org/10.1017/CBO9781316161012>.
- Walther L. Fischer, „Mathematik und Literaturtheorie“, in: *Archimedes* 22 (1970), 19–26.
- Fischer, Walther L., „Mathematische Texttheorie“, in: Heinz L. Arnold/Volker Sinemus (Hg.), *Grundzüge der Literatur- und Sprachwissenschaft*, Bd. 1: *Literaturwissenschaft*, München 61980, 44–61.
- Fitzpatrick, Kathleen, *Planned obsolescence. Publishing, technology, and the future of the academy*, New York u. a. 2011.
- Fitzpatrick, Kathleen, „Peer Review“, in: Susan Schreibman/Ray Siemens/John Unsworth (Hg.), *A New Companion to Digital Humanities*, Chichester, West Sussex/Malden, MA/Oxford 2016, 439–448.
- Greetham, David C., *Textual scholarship. An introduction*, New York u. a. 1994.
- Halbwachs, Maurice, *Das Gedächtnis und seine sozialen Bedingungen*, Berlin u. a. 1966.
- Hookway, Branden, *Interface*, Cambridge, MA, u. a. 2014.

¹⁰¹ Sämtliche digitalen Referenzen wurden letztmalig am 21.06.2018 eingesehen.

- Jannidis, Fotis/Kohle, Hubertus/Rehbein, Malte, *Digital Humanities. Eine Einführung*, Stuttgart 2017.
- Jochum, Uwe, „Die Aufgabe des höheren Dienstes“, in: Ders. (Hg.), *Der Ort der Bücher. Festschrift für Joachim Stoltzenburg zum 75. Geburtstag*, Konstanz 1996, 69–79.
- Jochum, Uwe, *Die Idole der Bibliothekare*, Würzburg 1995.
- Kammer, Stephan/Lüdeke, Roger (Hg.), *Texte zur Theorie der Texte*, Stuttgart 2005.
- Landow, George P., *Hypertext 3.0. Critical Theory and New Media in an Era of Globalization*, Baltimore 2006.
- Lankes, R. David, *The Atlas of New Librarianship*, Cambridge, MA, u. a. 2016.
- Leyh, Georg/Milkau, Fritz/Christ, Karl u. a. (Hg.), *Handbuch der Bibliothekswissenschaft*, 3 Bde., Wiesbaden 1931–1965.
- Limpinsel, Mirco, „Was bedeutet die Digitalisierung für den Gegenstand der Literaturwissenschaft?“, in: *Zeitschrift für digitale Geisteswissenschaften* 2016, DOI https://doi.org/10.17175/2016_009.
- Lorenz, Bernd, *Systematische Aufstellung in Vergangenheit und Gegenwart*, Wiesbaden 2003.
- Martínez, Matías, „Dialogizität, Intertextualität, Gedächtnis“, in: Heinz L. Arnold/Heinrich Detering, *Grundzüge der Literaturwissenschaft*, München 42001, 430–445.
- McCarty, Willard, *Humanities computing*, Basingstoke, Hampshire [u. a.] 2005.
- McGann, Jerome, *A New Republic of Letters. Memory and Scholarship in the Age of Digital Reproduction*, Cambridge, MA/London 2014.
- McLuhan, Marshall, *Understanding Media*, London u. a. 1964.
- Morhof, Daniel G., *Polyhistor sive de notitia auctorum et rerum commentarii*, Lübeck 1688.
- Mosco, Vincent, *The Digital Sublime. Myth, Power, and Cyberspace*, Cambridge, MA, 2005.
- Oehlmann, Doina, „Lizenzen oder Texte, Nutzung oder Hosting? Können Bibliotheken ihren Auftrag in Zeiten elektronischer Texte weiterhin ausreichend erfüllen?“, in: *ZfBB* 59 (2012) 5, 231–235.
- Osten, Manfred, *Das geraubte Gedächtnis. Digitale Systeme und die Zerstörung der Erinnerungskultur*, Frankfurt a. M. u. a., 2004.
- Pariser, Eli, *The filter bubble. What the Internet is hiding from you*, London 2011.
- Pédaquaque, Roger T., *Le document à la lumière du numérique. Présenté par Jean-Michel Salaün*, Caen 2006.
- Puhl, Johanna/Andorfer, Peter/Höckendorff, Mareike u. a., *Diskussion und Definition eines Research Data LifeCycle für die digitalen Geisteswissenschaften. DARIAH-DE Working Papers 11*, Göttingen 2015, <http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2015-11.pdf>.
- Raabe, Paul, *Ein Schatzhaus voller Bücher. Die Herzog August Bibliothek in Wolfenbüttel*, Hannover 31971.
- Raabe, Paul, *Die Bibliothek als humane Anstalt betrachtet*, Stuttgart 1986.
- Rusterholz, Peter, „Zum Verhältnis von Hermeneutik und neueren anti-hermeneutischen Strömungen“, in: Heinz L. Arnold/Heinrich Detering, *Grundzüge der Literaturwissenschaft*, München 42001, 157–177.
- Sahle, Patrick, *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, 3 Bde, Norderstedt 2013.
- Schnapp, Jeffrey T./Battles, Matthew, *The Library Beyond the Book*, Cambridge, MA, u.a. 2014.
- Schöch Christof, „Aufbau von Datensammlungen“, in: Fotis Jannidis/Hubertus Kohle/Malte Rehbein, *Digital Humanities. Eine Einführung*, Stuttgart 2017, 223–233.
- Schwarz-Friesel, Monika/Consten, Manfred, *Einführung in die Textlinguistik*, Darmstadt 2014.
- Stäcker, Thomas, „Ars sine scientia nihil est – Bibliothekswissenschaft aus forschungsbibliothekarischer Perspektive“, in: *Bibliothekswissenschaft – quo vadis? Eine Disziplin zwischen Tradition und Visionen. Programme – Modelle – Forschungsaufgaben*, München 2005, 33–46.
- Stäcker, Thomas, „Von Alexandria lernen – die Forschungsbibliothek als Ort digitaler Philologie“, in: *Frauen – Bücher – Höfe. Wissen und Sammeln vor 1800. Essays in honor of Jill Bepler*, Wiesbaden 2018, 93–104.

- Stäcker, Thomas, „Konversion des kulturellen Erbes für die Forschung. Volltextbeschaffung und -bereitstellung als Aufgabe der Bibliotheken“, in: *o-bib* 1 (2014), 220–237, DOI <https://doi.org/10.5282/o-bib/2014H1S220-237>.
- Stäcker, Thomas, „Vom Buch zum Text. Sammeln, Erschließen und Benutzen im digitalen Zeitalter“, in: *Buch – Bibliothek – Region. Wolfgang Schmitz zum 65. Geburtstag*, Wiesbaden 2014, 353–356.
- Wegmann, Nikolaus, *Bücherylabyrinth. Suchen und Finden im alexandrinischen Zeitalter*, Köln u. a. 2000.
- Werle, Dirk, „Literaturtheorie als Bibliothekstheorie“, in: Stefan Alkner/Achim Hölter (Hg.), *Literaturwissenschaft und Bibliotheken*, Göttingen 2015, 13–26.
- Winko, Simone, „Literarische Wertung und Kanonbildung“, in: Heinz L. Arnold/Heinrich Detering, *Grundzüge der Literaturwissenschaft*, München 42001, 585–600.
- Wirth, Sabine, „Between Interactivity, Control, and ‚Everydayness‘ – Towards a Theory of User Interfaces“, in: Florian Hadler/Joachim Haupt (Hg.), *Interface Critique*, Berlin 2016, 17–35.
- World Wide Web Consortium, *Extensible Markup Language (XML)* (8. Dezember 1997), <https://www.w3.org/TR/PR-xml-971208.html>.
- World Wide Web Consortium, *Resource Description Framework (RDF)* (25. Februar 2014), <https://www.w3.org/RDF/>.

Online-Ressourcen

- Datahub*, „Linked Open Emblem Data“, <https://old.datahub.io/dataset/emblem>.
- Deutsches Textarchiv (DTA)*, <http://www.deustextarchiv.de/>.
- Die Digitale Bibliothek bei TextGrid*, <https://textgrid.de/digitale-bibliothek>.
- Herzog August Bibliothek Wolfenbüttel. Forschungsprofil*, <http://www.hab.de/de/home/wissenschaft/forschungsprofil-und-projekte/forschungsprofil.html>.
- Nestor*, www.langzeitarchivierung.de.
- OCR-D*, www.ocr-d.de.
- Resource Description Access*, <http://access.rdatoolkit.org/> (Lizenz erforderlich).
- Text Encoding Initiative (TEI)*, <http://www.tei-c.org>.
- The Unicode Consortium*, <http://www.unicode.org>.
- The Unicode® Standard. Version 12.0 – Core Specification* (2017), www.unicode.org/versions/Unicode10.0.0/ch02.pdf.
- Verzeichnis der im deutschen Sprachbereich erschienenen Drucke des 16. Jahrhunderts (VD 16)*, www.vd16.de.
- Verzeichnis der im deutschen Sprachbereich erschienenen Drucke des 17. Jahrhunderts (VD 17)*, www.vd17.de.
- Verzeichnis der im deutschen Sprachbereich erschienenen Drucke des 18. Jahrhunderts (VD 18)*, www.vd18.de.
- Wikipedia*, „Semantic Web Stack“, https://en.wikipedia.org/wiki/Semantic_Web_Stack.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





'Could be the translation, of course'. Analysing the perception of literary fiction and literary translations

Karina van Dalen-Oskam

1 Introduction

The advent of information technology has greatly expanded the methodological toolkit of literary scholars. In the digital humanities project *The Riddle of Literary Quality*, methods from literary studies, linguistics and computer science are combined with sociological approaches. I will describe how this project went about mixing and blending methods from different research disciplines in search of more knowledge about the modern-day Dutch and European literary systems. My special focus is on translations of modern fiction and how they are valued by readers in comparison with fiction written in their native language. I will conclude by discussing how the greatly changed content of the toolbox may influence the planning and execution of a literary research project and what this could imply for the future of literary studies in general.

The Riddle of Literary Quality (2012–2019) combines a computational analysis of low-level (i.e. easy to measure) linguistic features and high-level (i.e. much more difficult to measure) linguistic features¹ with the results of a reader survey, in order to unlock part of the mystery that surrounds 'literary quality' as experienced by both general and specialist readers. The aim is to uncover what the current conventions of literariness are in the Netherlands, predominantly using methods that enable verification and replication of the results. In addition to myself as the principal investigator/project leader and temporary collaborators from various

¹E.g. syntax; Andreas van Cranenburgh, *Rich statistical parsing and literary language*, Amsterdam 2016, <http://hdl.handle.net/11245/1.543163> (retrieved on June 11, 2019).

K. van Dalen-Oskam (✉)
Huygens Institute, Amsterdam, Netherlands
E-Mail: karina.van.dalen@huygens.knaw.nl

areas, in the earlier phase of the project (2011–2016) the team also comprised two PhD students and a scientific programmer funded by the *Royal Netherlands Academy of Arts and Sciences (KNAW)* in its *Computational Humanities Programme* (2011–2016). The project is now in its final phase.²

The corpus that we selected for our research contains the 401 best-selling or most often borrowed (from public libraries) novels in the period 2010–2012 in the Netherlands, in order to ensure that we could gather enough opinions per book for statistical observations. As we wanted to exclude as much as possible the influence of long-standing ideas about which titles belong to the literary canon and which do not, we selected novels that were generally not published in Dutch before 2007. Our selection criteria led to a corpus that consisted not only of high-brow literary novels, but also of many popular novels that until recently were not often considered as objects of research by students of literature, such as fiction for young adults and suspense, romance and fantasy novels. Another unusual choice was to include novels translated into Dutch, as long as they matched the selection criteria. Translations play a very important role in the Netherlands. In our corpus, 152 titles were written in Dutch and 249 were translated into Dutch and published in Dutch for the first time in 2007–2012. Of these 249 novels, 180 were translated from English. The remaining 69 were translated from nine other languages. One of the project's research questions was whether there are any differences in how Dutch originals are valued on the scale of literariness compared with translations into Dutch.

Readers opinions were gathered in a large online survey, the *National Reader Survey (Het Nationale Lezersonderzoek)*, which ran from March to September 2013. Almost 14,000 respondents completed the survey. They provided some personal information and answered questions relating to their reading habits, and indicated which of the 401 novels they had read. We then presented a randomly selected list of seven of the books they had read and asked them how good they found each of these books on a scale of 1 (very bad) to 7 (very good), and how literary they found the book on a scale of 1 (not literary at all) to 7 (highly literary). Respondents were then offered the option to rate seven more titles. For one of the books, we asked them to account for their score on the scale of literariness in the form of a written motivation. For seven books a respondent had not read, we asked for a comparable rating. We did not ask them to motivate their score for a book they did not read.

²For an introduction to the project, see Karina van Dalen-Oskam, "Prehistory of The Riddle" ("The Riddle of Literary Quality: The search for conventions of literariness," transl. of: "The Riddle of Literary Quality. Op zoek naar conventies van literariteit," in: *Vooy's: tijdschrift voor letteren* 32/3 [2014], 25–33), <http://literaryquality.huygens.knaw.nl/?p=537#more-537> (retrieved on June 11, 2019).

2 Research Questions

Dutch literature in its international context is the topic of a recently published volume titled *Doing Double Dutch*,³ which focuses on how Dutch literature has been and is being transferred to other language areas through translations and adaptations. Translators are key players in this kind of cultural transfer and several chapters deal with specific literary works and how they have been translated into different languages and cultures using slightly different approaches.

In the Netherlands, 65–70 per cent of all fiction titles are translations,⁴ mostly from English. But English-language areas are not the most important receivers of translations from Dutch. Most scholars contributing to *Doing Double Dutch* agree that it is the German language area that is most important and often also the means by which interest in Dutch literary works is generated in yet other language areas. It is well known that interest in other literatures is neither mutual nor mirrored, and that imbalance is one of its key characteristics. There is a clear hierarchy in the 'world republic of letters',⁵ wherein English currently occupies the centre and Dutch literature is one of the many literatures on the periphery. The contributors to *Doing Double Dutch* repeatedly state, however, that cultural transfer can only take place when the periphery is somehow present in the centre, and vice versa. The methods they use are mainly qualitative by nature, and they do not make use of interviews, surveys, statistical approaches or computational stylistics (*stylometry*)—all methods that play a role in *The Riddle of Literary Quality*.

Against this backdrop I will first focus on Dutch readers. Using the results of the *National Reader Survey*, I will address the following three research questions:

1. How aware are Dutch readers that they are reading a translation?
2. Do Dutch readers give translated novels different scores compared with novels written in Dutch?
3. Do Dutch readers use different criteria when accounting for their appreciation of translated fiction compared with fiction written in Dutch?

The answers to these three questions lead to a new question: do we see the same patterns in other countries, or are there differences? To answer that I will shift my focus to a larger geographical area and describe an international experiment involving the Netherlands, Germany, France and the United Kingdom and summarize its preliminary findings. These findings will lead back to *The Riddle of*

³Elke Brems/Orsolya Réthelyi/Ton van Kalmthout (eds), *Doing Double Dutch: The international circulation of literature from the Low Countries*, Leuven 2017.

⁴Thomas Franssen, *How books travel. Translation flows and practices of Dutch acquiring editors and New York literary scouts, 1980–2009*, Amsterdam 2015, 30, <https://hdl.handle.net/11245/1.446961> (retrieved on June 11, 2019).

⁵Pascale Casanova, *The world republic of letters*, translated by M.B. DeBevoise, Cambridge, MA/London 2004.

Literary Quality and a fourth research question, which will for now be addressed only very summarily:

4. What does a stylometric comparison of translated novels and novels written in Dutch reveal?

3 The *National Reader Survey*

The almost 14,000 respondents who completed the *National Reader Survey* are representative of the Dutch reading audience. Just over 70 per cent of the respondents are female and about the same percentage have a high level of education (university or higher vocational education); however, the men on average are slightly higher educated than the women. These are the people who read, buy novels, join reading groups and make use of libraries. The respondents also include professional readers, such as reviewers. We know this because they sometimes mention this in the open field that asks them to account for one of their ratings.

3.1 Translation Awareness

How aware are Dutch readers that they are reading a translation? Participants in the *National Reader Survey* were asked to account for their rating on the scale of literariness of a randomly selected book to which they had given a relatively high or a relatively low score in comparison with the other books they had rated. Some of the explanations they gave suggest that readers are very aware of whether they are reading a translation. Some respondents indicated they had read a book in the original language:

Heb het in het Frans gelezen. Prachtig taalgebruik en actuele problematiek zeer origineel en tussen de regels weergegeven.

Read it in French. Beautiful language use, current issues, very original and 'between the lines'.

Personen worden nogal vlak neergezet. Stijl is matig. (In het Engels gelezen) Veel herhalingen.

Characters are rather flat. Style not very good (read it in English). A lot of repetition.

Respondents probably often gave their opinion of a book based on the original and without having read the Dutch translation. We do not think this is problematic for our survey—we were happy with an opinion even when it was based on the original. Some respondents addressed this issue explicitly, commenting on the quality of our survey:

Ik mis de vraag of ik het in het Engels heb gelezen: ja Julian Barnes schrijft fraaie, korte zinnen. De opbouw van al zijn boeken is systematisch de verhaallijn(en) is/zijn duidelijk. Zijn Engels is zeer goed te begrijpen in tegenstelling tot bijv. Coetzee die veel 'gemaakter' schrijft. (...)

You did not ask if I have read it in English: yes. Julian Barnes writes beautiful, short sentences. All his books are well structured and the plot line or lines is or are clear. His English is easy to understand, compared with for instance Coetzee, who has a much more 'artificial' style. (...)

Literair zegt voor mij iets over het taalgebruik (daarom heb ik dit boek dan ook in het Engels gelezen). Soms is een boek in de oorspronkelijke taal voor mij literair, terwijl de Nederlandse vertaling dat absoluut niet is!! Slechte vertaling of matige vertaling kan van een literair boek een flodder maken. Dat betekent dus ook dat ik uw insteek van deze enquête niet goed vind. Ik wil namelijk niet een vertaler op zijn literaire waarde beoordelen, maar wel de schrijver. En ik weet van veel anderen dat zij zich ook vaak ergeren aan slecht vertaalde boeken.

To me, 'literary' has to do with language use (that's why I read this book in English). Sometimes, I find a book literary in its original language, while the Dutch translation is definitely not literary! A bad translation or one of moderate quality can turn a literary book into a mess. This also means that I don't approve of your approach in this survey. I want to evaluate the literary quality of the author, not of the translator. And I know that many other people are also annoyed by badly translated books.

Respondents praised a translation only occasionally. More often, they tried to soften their bad opinion of a book by adding a phrase such as 'could be the translation, of course'.

Te oppervlakkig. Gaat niet diep genoeg op de materie in. Kan ook aan de vertaling liggen.

Too shallow. Does not delve deep enough into the topic. Could also be the translation.

Verhaal een beetje over de top. Niet altijd mooi van taal. (Kan ook aan de vertaling liggen natuurlijk!)

Story a bit over the top. Language not always agreeable. (Could also be the translation, of course!)

3.2 Differences in Ratings

Do Dutch readers give translated novels different scores compared with novels written in Dutch? I will only go into the scores on the scale of literariness. For all 401 novels, the mean score was 4.67 on a scale of 1 (not literary at all) to 7 (highly literary). Splitting the corpus into two revealed an interesting result: the mean score on the scale of literariness for all 152 Dutch novels is 4.95, whereas for all 249 novels translated into Dutch it is 4.47.

We need to look at different genres, however. There are 147 literary novels in the corpus—novels published with the label 'literary novel' given to it by the publisher and by libraries. The most-represented genre fiction in our corpus is the suspense novel (186), followed by the romance novel (41). In a fourth category, 'Other' (27), we put genres with only a couple of representatives, such as fantasy and regional novels, and titles that were marketed as novels but may also be

Table 1 Number of titles in the corpus per genre

	Dutch titles	Translated titles	Total
Literary novel	73	74	147
Suspense	58	128	186
Romance	5	36	41
<i>Other</i>	<i>16</i>	<i>11</i>	27
Total	152	249	401

Table 2 Mean score on the scale of literariness per genre

	Dutch titles	Translated titles	Total
Literary novel	5.52	5.35	5.21
Suspense	3.76	4.02	3.94
Romance	2.75	2.97	2.94
<i>Other</i>	<i>4.74</i>	<i>3.54</i>	4.25
Total	4.95	4.47	4.53

labelled as literary non-fiction. We also put several volumes of short stories in this category. These 27 titles are thus a mix of different kinds of texts and are excluded from the comparison because there are too few of them to yield trustworthy statistics. Table 1 shows the number of titles in each of the categories, and Table 2 presents the mean scores on the scale of literariness.

What we see here is that the mean score on the scale of literariness for Dutch titles in the category of the literary novel is higher than for translated titles, whereas the opposite is found for the suspense novel and the romance novel, where the mean scores for translated titles were higher than for Dutch titles.

3.3 Differences in Criteria

Do Dutch readers use different criteria when accounting for their appreciation of translated fiction compared with fiction written in Dutch? Here, the explanations given by the respondents functioned as a proxy, since they do not provide the explicit criteria that the respondents used. But their explanations can perhaps help us to come a bit closer to understanding their criteria, whether applied consciously or not. I made use of *AntConc*, the freely available concordancing software developed by Laurence Anthony.⁶

I created two files, one containing all explanations relating to the 152 Dutch titles and the other with all explanations related to the 249 translated books. I then compared the vocabularies of both files, first listing the words that in a relative sense occur far more often than expected in the explanations related to Dutch titles than in those related to translated books, then listing the words that are relatively

⁶Laurence Anthony, *AntConc* (Version 3.4.3), Tokyo 2014, <http://www.laurenceanthony.net/> (retrieved on June 11, 2019).

Table 3 Keywords (nouns) in explanations related to Dutch titles compared with translated titles (3a) and keywords in explanations related to translated titles compared with Dutch titles (3b)

3 A	Keywords in explanations related to Dutch titles	3B	Keywords in explanations related to translated titles
<i>afstand</i>	distance	<i>betekenis</i>	meaning
<i>boekje</i>	booklet	<i>boek</i>	Book
<i>compositie</i>	composition	<i>boeken</i>	books
<i>gebruik</i>	use	<i>chicklit</i>	chick-lit
<i>jeugd</i>	youth	<i>deel</i>	volume
<i>journalistiek</i>	journalism	<i>fantasy</i>	fantasy
<i>moeder</i>	mother	<i>genre</i>	genre
<i>proza</i>	prose	<i>herhalingen</i>	repetitions
<i>stijl</i>	style	<i>hoofdpersonen</i>	main characters
<i>taal</i>	language	<i>naam</i>	name
<i>taalgebruik</i>	language use	<i>serie</i>	series
<i>vader</i>	father	<i>thriller</i>	thriller
<i>verbeelding</i>	imagination	<i>thrillers</i>	thrillers
<i>woordgebruik</i>	word use	<i>verhaallijnen</i>	plot lines
<i>zinnen</i>	sentences	<i>verstand</i>	intellect
<i>zoon</i>	son	<i>vertaling</i>	translation

far more frequent in the explanations related to translated titles compared with those related to the Dutch books. For this, I used the keyness function in *AntConc*, which is based on log likelihood. In describing the results, I focus on the appellative nouns with the highest keyness, calling these ‘keywords’. Table 3 shows that the list of top key nouns in explanations related to the scores on the scale of literariness of Dutch novels contain several words referring to writing style and language use, and words referring to close family members. The top key nouns in the explanations related to translated novels mainly refer to genres, plot and serialization, and also include the word ‘translation’.

The *Riddle* corpus contains an almost equal number of literary novels written in Dutch (73) and literary novels translated into Dutch (74). This means that the number of translated genre fiction titles is much higher than the number of original Dutch genre fiction titles, and this will certainly have influenced the results in Table 3. I therefore selected only the explanations related to the category of literary fiction and compared the vocabularies in the explanations related to Dutch literary novels with those related to translated literary novels. This led to Table 4, which again shows the nouns with the highest keyness in each of these files.

Compared to Table 3 and 4 shows a partly different picture, especially regarding the explanations related to the translated literary novels, where genre

Table 4 Keywords (nouns) in explanations related to Dutch literary titles compared with translated literary titles (4a) and keywords in explanations related to translated literary titles compared with Dutch literary titles (4b)

4 A	Keywords in explanations related to Dutch literary titles	4B	Keywords in explanations related to translated literary titles
<i>taalgebruik</i>	language use	<i>vertaling</i>	translation
<i>moeder</i>	mother	<i>verhaallijn</i>	plot
<i>taal</i>	language	<i>herhalingen</i>	repetitions
<i>journalistiek</i>	journalism	<i>verhaal</i>	story
<i>proza</i>	prose	<i>deel</i>	volume
<i>jeugd</i>	youth	<i>boek</i>	book
<i>stijl</i>	style	<i>hoofdpersonen</i>	main characters
<i>boekje</i>	booklet	<i>spel</i>	play
<i>uitdieping</i>	in-depth study	<i>kaart</i>	map
<i>verdriet</i>	sorrow	<i>verhouding</i>	relation
<i>kinderen</i>	children	<i>drijfveren</i>	motives
<i>verhaaltjes</i>	stories	<i>huwelijk</i>	marriage
<i>zinnen</i>	sentences	<i>richting</i>	direction
<i>afstand</i>	distance	<i>slavernij</i>	slavery
<i>poes</i>	cat	<i>woordkeus</i>	word choice
<i>volzinnen</i>	stylish sentences	<i>kind</i>	child

names do not reappear. However, the keywords related to the Dutch literary novels still show slightly more nouns associated with style and language use. The keywords in the explanations related to translated literary fiction may be characterized as focusing on narrative structure and themes in the story.

4 A New Experiment

The answers to the first three research questions suggest that when asked for their opinion about a novel's literariness, Dutch readers see Dutch novels in the category 'literary novel' as slightly more literary than translated novels in the same category. Both Dutch suspense novels and romance novels, however, get a slightly lower mean score on the scale of literariness than translated novels in these categories. In accounting for their scores, Dutch readers more often referred to language use and style (and family relations) when dealing with Dutch titles than when they accounted for their opinions of translated books. These Dutch readers thus seem to be inclined to see their 'own' literature as somewhat more literary than translated literature. Do we see the same pattern in other countries?

The third chapter in *Doing Double Dutch* (which was not yet available when we developed the experiment I will report on below) deals with a small part of this topic. Nico Wilterdink gives a very short overview of the international reception of contemporary Dutch literature in the United States, the United Kingdom and France, and he pays special attention to the historically changing literary relation with Germany. Wilterdink views the literary world-system as consisting of nationally or linguistically bounded literary fields that are interconnected through asymmetrical relations of power and influence. Worldwide, more than 50 per cent of translations are from English, but only 4 per cent of all books published in English are translations. Wilterdink observes: "Apart from some academic circles, there is hardly any public concern about a lack of openness to foreign (i.e. originally non-English) literature, if this is observed at all."⁷ For completeness' sake: in Germany and France, the share of translations is 14–18 per cent.⁸

Wilterdink notices a growing interest in Dutch literature since the 1980s for the countries he deals with, an observation he based on a growing number of translations and reviews of these translations.⁹ He draws attention to the fact that some Dutch authors are more famous in Germany than in the Netherlands. And in France, he states, the Netherlands (from a literary perspective) is seen as a country that is far, far away in the north—part of the Scandinavian countries. In the United Kingdom, translated literature from the Netherlands is even less visible than in France. The British do not seem to like books by authors with funny names, and Wilterdink assumes widely divergent stylistic conventions in the UK and on the European mainland may also explain this invisibility.

If we summarize the observed differences between the four countries in a few catchwords, we may say that the American attitude to Dutch literature is predominantly one of indifference and ignorance outside a small circle of professional specialists; the English attitude a mixture of arrogance and self-criticism; the French attitude well-meaning but superficial; and the German orientation open, interested, often enthusiastic.¹⁰

Wilterdink tries to explain this situation by looking at the global literary field or world-system as part of the more encompassing cultural world-system:

In general terms, we can say that the geographical, social and cultural distance between two nations correlates inversely with the intensity of transnational cultural exchange: the larger the distance, the less cultural exchange. Among the four nations compared,

⁷Nico Wilterdink, "Breaching the dyke. The international reception of contemporary Dutch translated literature," in: Elke Brems/Orsolya Réthelyi/Ton van Kalmthout (eds), *Doing Double Dutch: The international circulation of literature from the Low Countries*, Leuven 2017, 45–65, here: 54.

⁸Cf. Franssen (Ann. 4), p. 89, for more information.

⁹Wilterdink (Ann. 7), p. 51.

¹⁰*Ibid.*, 53.

the distance with the Netherlands is the largest for the United States, the smallest for Germany, with France and England in between; and the differences in attention to Dutch literature correspond with this variation.¹¹

For our research, we wanted to look beyond only translated Dutch literature. We especially wanted to find out whether readers in other countries or language areas also look differently at novels written in their own language as compared with novels translated from other languages. We decided to perform an experiment in which we focused on novels from the literary category, excluding suspense and romance novels to not overburden the survey with too many variables. Our research questions for this experiment were:

- a) Do respondents' scores on the scale of literariness differ for novels written in their native language compared with novels translated into their native language?
- b) If there is a difference, is there a certain hierarchy in the countries or languages of origin of the translated book and, if so, is this ranking the same in different countries?

We hired an international panel of readers with backgrounds comparable to those of the participants in the *National Reader Survey*. The gross representative sample consisted of participants aged 18 to 65 years. The net number of participants per country was 150.

We aimed for an equal number of novels in each of the main languages of four to six European countries, with an equal number of novels written by female and male authors. Only novels that were available in all chosen languages were selected, which proved to be a very difficult criterion. We hoped we could use titles that were also in the corpus of *The Riddle of Literary Quality* and the *National Reader Survey*, but this was only partially possible. To my dismay, the most challenging part was finding enough titles written by women.

From Wilterdink's perspective it is probably not due to chance that we ended up with the same selection of countries as he did, only excluding the United States. We gathered a list of 24 novels, six from each of the following countries/language areas: the Netherlands (Dutch), the United Kingdom (English—we deliberately limited our selection of novels to authors from the British Isles), France (French) and the German language area (German—with titles published in Germany, but some of the authors originating from an adjacent country). Table 5 presents the complete list.

We asked the respondents to indicate which of these books they had read, and also to score all of them on the scale of literariness (instead of asking this only for the books they had read). Compared with the results of the *National Reader Survey*, this implied that most scores from the new experiment would very probably be

¹¹ *Ibid.*, 55.

Table 5 The corpus of 24 novels available in Dutch, English, French and German

Title	Author	Originally published in
<i>22 Britannia Road</i>	Amanda Hodgkinson	United Kingdom
<i>Contrapunt</i>	Anna Enquist	Netherlands
<i>Tirza</i>	Arnon Grunberg	Netherlands
<i>Melnitz</i>	Charles Lewinsky	Germany
<i>Feuchtgebiete</i>	Charlotte Roche	Germany
<i>The Thousand Autumns of Jacob de Zoet</i>	David Mitchell	United Kingdom
<i>Rien ne s'oppose à la nuit</i>	Delphine de Vigan	France
<i>Room</i>	Emma Donoghue	United Kingdom
<i>De Omweg</i>	Gerbrand Bakker	Netherlands
<i>Het Diner</i>	Herman Koch	Netherlands
<i>Atemschaukel</i>	Herta Müller	Germany
<i>Bring up the Bodies</i>	Hilary Mantel	United Kingdom
<i>Adam und Evelyn</i>	Ingo Schulze	Germany
<i>L'Africain</i>	J.M.G. Le Clézio	France
<i>The House of Special Purpose</i>	John Boyne	United Kingdom
<i>Die Mittagsfrau</i>	Julia Franck	Germany
<i>The Sense of an Ending</i>	Julian Barnes	United Kingdom
<i>HhhH</i>	Laurent Binet	France
<i>De Verdronkene</i>	Margriet de Moor	Netherlands
<i>Ladivine</i>	Marie Ndiaye	France
<i>Réparer les Vivants</i>	Maylis de Kerangal	France
<i>La Carte et le Territoire</i>	Michel Houellebecq	France
<i>Een Hart van Steen</i>	Renate Dorrestein	Netherlands
<i>Der Turm</i>	Uwe Tellkamp	Germany

primarily based on the reputation of a text or author (or language or national literary field) and not on the reader's actual experience of the text itself. We are, of course, aware that even when people read a book, reputation also plays a role. But at least in theory, when they did read a book, their score may also be informed by the actual reading experience and thus also by stylistic, narrative, etc. elements.

We used the same 7-point scale as in the *National Reader Survey* and presented the 24 novels in random order. In the Netherlands, we presented the Dutch titles of the translated novels, followed by the author's name and the country in which the book was originally published. In the United Kingdom, the first column had all titles in English, in France all titles in French and in Germany all titles in German. I will not describe all questions and all results of this experiment, but focus on answering the two research questions presented above. Statistical details will also be published elsewhere.

4.1 Differences in Ratings

Do respondents' scores on the scale of literariness differ for novels written in their native language compared with novels translated into their native language? Yes, there is a difference. The *National Reader Survey* results showed that Dutch readers on average gave a slightly higher score on the scale of literariness to novels written in Dutch than to novels translated into Dutch. In the international panel survey, we see the same results for the Netherlands, the United Kingdom and France. Germany is an exception, in two ways. One of these relates to the novel *Feuchtgebiete* by Charlotte Roche. In Germany, this book received a (statistically significant) lower score than in the three other countries where it received a (statistically significant) higher score. The German panel was also exceptional in another respect. Compared with the panels in the other three countries, the representative panels in France, the United Kingdom and the Netherlands gave significantly higher scores to only two of their 'own' novels, and to *Feuchtgebiete*. The German panel had only one title with a significantly higher score for books from *each* of the four countries. For the German language area this was *Der Turm* by Uwe Tellkamp, for English *The Sense of an Ending* by Julian Barnes, for French *Réparer les Vivants* by Maylis de Kerangal, and for Dutch *De Verdronkene* by Margriet de Moor. The German panel seems to be more fair and more open to translated novels than the other panels, which seems to accord with the observations by Wilterdink in his chapter in *Doing Double Dutch*. Of course this is a very preliminary conclusion and it needs to be followed up with more in-depth research. It could, for instance, also be the case that national and international literary prizes had a larger impact on the German panel members than on the members of the other panels.

4.2 A Literary Hierarchy of Countries

Apart from the literary appreciation scores, we also have the answers to a couple of other questions we asked the representative panels in the four countries. After they had scored the 24 novels, we asked them how they valued the literatures of the different languages in comparison with each other. They were not allowed to go back in the survey to see the scores they had given to the 24 novels. The German panel, for example, was asked (in German):

If you compare English literature to German literature, which of the following statements do you find most adequate?

1. There is no difference.
2. English literature usually has a higher literary quality.
3. German literature usually has a higher literary quality.
4. I do not know.

Table 6 Respondents' ranking of the literariness of the literatures of all four countries

United Kingdom	France	Netherlands	Germany
1. English	1. French	1. English	1. German
2. French	2. English	2. Dutch	2. English (close behind German!)
3. German	3. German	3. French	3. French
4. Dutch	4. Dutch	4. German	4. Dutch

The same question was then asked for French compared with German and Dutch compared with German. The panels in the other countries were asked to make the same comparisons of the other languages with their own language. The order of the three questions about the three languages was always random. Next, the panel members in all countries were asked to rank the four languages, putting the language with the highest literary quality first and that with the lowest literary quality last.

These questions also yielded very interesting results. For the first set of three, many respondents selected the first option: there is no difference. But there were also enough respondents opting for another choice. Taking all responses into account, we can summarize that the English panel had a higher regard for English literature than for Dutch and German literature, and sees it as comparable to French literature. The French nicely mirrored this: the French panel had a higher regard for French literature than for Dutch and German literature, and saw it as equal to English literature. The Dutch panel opined that Dutch literature did not reach the high level of literariness of English literature but was certainly comparable to French and German literature. And finally, the German panel indicated that German literature did not reach the level of English literature but was comparable to French literature, and was definitely more literary than Dutch works. The respondents' ranking of all four national literary fields is presented in Table 6.

It will be clear that according to the members of the four national panels, there is a hierarchy. It is also clear that this hierarchy differs in each country. In Germany, German literature is valued higher than literature from the other three countries, although only just beating English. In France, French literature is valued most and in the United Kingdom English literature is valued most. Only the Dutch panel sees its 'own' literature as second to another one, namely English literature. But this is not true for the scores the Dutch panel gave—there, the panel members (unconsciously?) put Dutch literature first. And this was also the outcome of the *National Reader Survey*.

The German panel's scores for the 24 novels also do not quite agree with the results of the explicit rankings. This definitely needs more research, but for now, I expect that the mean of the ratings of all the books may give a better idea of the situation than the explicit rankings. In comparing or ranking the four literatures,

Table 7 Combined ranking of the literariness of the literatures of all four countries

1. English
2. French
3. German
4. Dutch

respondents may quite easily answer in a more politically correct way, whereas their individual scores may bring to light any unconscious, implicit biases they have.

If we combine the rankings of all panels, we see a result that seems to agree with current ideas about the relative place of the four languages in the hierarchy of the world republic of letters (Table 7).

After the respondents had presented their ranking of the four literatures, we asked them if they could explain why they considered novels from their chosen first rank to be the most literary. We have 469 responses to this question, some of which only state that they cannot explain their choice, that it is just a feeling or just their own opinion. Others refer to the style, or that they know the language and the culture and thus cannot make a fair comparison. Deeper layers as well as plot lines are also mentioned.

It is rather difficult to do a vocabulary analysis of these answers, since they are in four different languages. A word frequency list also does not help that much. It does help, however, to provide a quick overview of all the words used, to quickly find for instance all personal names that are mentioned: William Shakespeare (9 times), Jane Austen, Charles Dickens (4 times), Harry Potter (3 times), J.K. Rowling, Brontë sisters (twice each), Arnon Grunberg, Honoré de Balzac, Julian Barnes, Brecht, Charlotte Brontë, Camus, Eliot, Jane Eyre, Goethe, Hesse, Mann, Terry Pratchett, George Orwell, Siegfried Sassoon, Thackeray and J.R.R. Tolkien (once each). Below I present some of the explanations given by respondents favouring English literature, to illustrate the kind of answers we received:

Aus dem gefühl heraus

Just a feeling

Based on my own nationality

Beaucoup d'auteurs littéraires

Many literary authors

Beter woordgebruik en andere thema's

Better language use and different themes

Haben meist tolle handlungen und schauplätze

Usually have great plots and settings

Harry Potter is the best thing ever

Hebben bij mijn weten de meeste grote auteurs voortgebracht

To my knowledge have the largest number of great authors

Het is gewoon een feit, net als britse films of series

It's just a fact, same as British films or series

We are the home of it :)

We have Dickens and Eliot etc.

We have the best authors

We have the most famous writers in history and we have learnt from them.

And some examples from those favouring German literature:

Allein durch die Namensangabe
 Just by looking at the names
 Anything is better than English
Ben de taal magtig [sic: spelling error for 'machtig']
 I master the language
 Better arcing of plot lines.
Das land der dichter und denker
 The country of poets and thinkers
Eigentlich kommt es nicht darauf an aus welchem Land ein Buch kommt. Es gibt in jedem Land bessere und schlechtere Autoren. So wie bei allen Sachen...
 In fact it doesn't matter which country a book comes from. Each country has good and bad authors, as with anything...
Finde ich am besten geschrieben
 I think they are the best written
Plus détaillé
 More detailed

I will not go deeper into this here. A more thorough analysis will be part of another publication.

5 Stylometry

As shown above, direct questions to readers may not provide us with the best answers to our research questions. But a broader perspective, combining scores and explanations, seems more useful because then we can actually see that respondents are not always aware of any biases they have. The scores of the novels in the new international experiment showed that representative panels of readers from three of the four countries tended to rank literature written in their own language as the most literary. This agrees with the results of the *National Reader Survey*.

The next question is then: why? Can it be that some or all of the actual translations are of a lower linguistic and literary quality than the originals, as some of the respondents in the *National Reader Survey* suggested in their explanations? Even if that were not the case, we may also consider a more sociological explanation. Perhaps readers are not inclined to consider a translation as a typical or prototypical literary novel. This may be because in the literature classes at the schools they attended as young adults they were trained to pay attention only to literature in the native language, which implicitly demoted the worth of translations. This could be researched by, for instance, looking at how the literary canon was and is taught in different countries. Could it be that in the Netherlands, the United Kingdom and France, schools promote/promoted a clear divide between national literature and such a thing as 'world literature' or the classics in the high school curriculum, and that this is/was different in Germany, to name but one possibility? I will not go into this here, but I did take a first step towards answering the question concerning translation quality by performing a small stylometric experiment.

Table 8 Literary novels

	Female author	Male author
Written in Dutch	23	50
Translated into Dutch	34	40
Translated from English	25	19
Total	57	90

What does a stylometric comparison of translated novels and novels written in Dutch reveal? I again used the keyness option in *AntConc* to address this question. This time, I compared the vocabularies of a subset of the literary novels written in Dutch with novels translated into Dutch to see whether this would give us any new points of connection for further research. In selecting the subset, the gender of the author also played a role. As PhD student Corina Koolen shows in her thesis about *The Riddle of Literary Quality*,¹² in the *National Reader Survey* novels written by male authors tended to be seen as more literary than those by female authors. For my small stylometric experiment, I therefore used only an all-male or an all-female authored subcorpus. My aim was to compare a roughly equal number of Dutch original literary novels with a set of novels translated from one language, all in the same category or genre. Table 8 shows that this almost automatically leads to the choice of literary novels written by female authors in Dutch and translations into Dutch from English novels written by women. The *Riddle* corpus has 23 original Dutch literary novels written by women and 25 translations from English novels authored by women.

Table 8 shows that of the 401 Dutch best-sold and most-borrowed titles, 50 novels written by male Dutch authors made it to the list, and only 23 by women did. Slightly more translated titles were written by male authors (40) than by female authors (34). The number of translations from English was higher for the translated titles of female authors (25) than male ones (19).

When we compare the vocabularies of all 23 novels written in Dutch by women with those of the 25 novels by women translated from English, the first thing that shows up is that most of the top 100 keywords are personal names. This is not unexpected. Proper names usually end up high on keyword lists, because they are mostly closely linked to the specific story that is being narrated and do not occur in many others. An analysis of the kinds of names in both sets is certainly interesting, but will have to wait for another occasion.

What is interesting in both top 100 lists is the occurrence of personal pronouns. In the keywords characteristic of original Dutch novels we find the following

¹²Corina Koolen, *Reading beyond the female: The relationship between perception of author gender and literary quality*, Amsterdam 2018, <https://hdl.handle.net/11245.1/cb936704-8215-4f47-9013-0d43d37f1ce7> (retrieved June 11, 2019).

pronouns: “ik”, “me”, “mij”, “mijn” (‘I’, ‘me’, ‘my’, ‘mine’). In the keywords for the translated novels, however, we find “ze”, “haar”, “hen” (‘she/they’, ‘her’, ‘them’). This seems to indicate a possible difference in perspectives: are the most popular Dutch novels by women perhaps more often first-person narrations, while the translated ones prefer a third-person approach? Or are the Dutch novels more ‘egocentric’ in some other way, for instance through a more prominent role of dialogue?

Another observation concerns nouns referring to close family members. The top 100 Dutch novels’ keywords include “mama” (‘mummy’), “oma” (‘granny’) and “vader” (‘father’), whereas the translated novels has “mam” (‘mum’) in the keywords. What especially interests me in this list is the contrast between “mama” and “mam”: Dutch authors give preference to “mama” and translations into Dutch more often use “mam”. This could be an indication that the word form “mam” is influenced by English “mum”, and thus that the relatively higher use in translations from English somehow shows the English-language origin and might say something about the nature (and perhaps even the fluency?) of the translation. One should remark, however, that some Dutch women prefer to be called “mam” and not “mamma”—which ofcourse in its turn may be interpreted as an indication of a higher status of words with an English connotation.

From a translation perspective, the next two observations are also very interesting. Among the top 100 keywords of the translated novels, we find a number of verbs that probably mostly function to indicate turns in dialogue: “vroeg” (‘asked’), “wist” (‘knew’), “zei” (‘said’). This could mean that turns in dialogue are usually more explicitly described in the English originals than is usual in Dutch novels, or it could mean that the translations did not quite represent it correctly in the Dutch target language—and thus could be a sign of a fluency issue in the translations. The same can be said for the following words with high keyness in the translated novels: “en” (‘and’), “toen” (‘then’), “dat” (‘that’), “voordat” (‘before’), “terwijl” (‘while’), “hoewel” (‘although’). These conjunctions occur many more times in the translations than in the originals, and this again could indicate where the translations did not quite match the style as we find it in the selected novels written in Dutch.

Without further analysis we cannot exclude the possibility that the above observations are indications of translations not quite being up to the standards that, in this experimental setup, the originals can be seen as providing. This definitely could be *part* of the answer to the question why Dutch readers value the literariness of Dutch literary novels slightly higher than translated ones. However, more in-depth research from a translation perspective is necessary, using the tools and methods of literary stylometry and corpus linguistics.¹³ We could, for instance, trace and analyse the usage of the words with high keyness

¹³ Cf. for a first exploration in this direction: Ton van der Wouden/Saskia Lensink/Karina van Dalen-Oskam, “Hidden indicators. Modal Particles in modern Dutch fiction and fiction translated from English: a quantitative approach” (forthcoming).

in the individual novels in the two subcorpora we compared, and see whether the individual translations differ in their use of these words and if any differences correlate with the mean ratings for the individual titles. Are some translations considered to be more literary than others? And are these the ones with a less deviating frequency of some of the words with keyness in the translation corpus? We should also bring in the English originals in digital form, for instance using an alignment tool, to make it easier to check the occurrences of the keywords in the translations against the original English phrases they were meant to represent. This may lead to plenty of interesting follow-up research, criss-crossing even more disciplinary boundaries than has already been done in this research project.

6 Conclusion

In seeking to uncover the current Dutch conventions of literariness, *the Riddle of Literary Quality* project has crossed several disciplinary borders. We have embraced many of the opportunities created by technological advancements. First, the Netherlands has one of the highest internet penetration rates in the world—in both 2011 and 2016, 94 per cent of all households had internet access, and in the latter year 92 per cent of all households actively made use of it.¹⁴ This made it possible to take the *National Reader Survey* online. Setting up and conducting the survey online is financially much more feasible than a manual approach. Using email and various kinds of blogs enabled us to find plenty of respondents in a time-efficient manner—we even had enough budget left to hire a public relations specialist to help us. It was also quite easy to gather the results and analyse them. We made use of a professional market researcher, Erika Nagelhout, and her national and international network, and this led to a very interesting and inspiring collaboration, resulting in several small follow-up experiments, including the international experiment described above. We think that this approach of finding respondents online and hiring representative panels makes our research and its results much more representative than many experiments done by psychologists whose test groups consist of first-year psychology students taking a survey for credits or money.

Second, the advent of the e-book came just in time. When I drafted the first version of the research plan in 2010, publishers in the Netherlands were still very hesitant to produce e-books. But by the time the proposal was reviewed and granted and we got ready to start in 2012, we were pleased to find that most of the 401 novels on our list were available in digital form. We only had to scan and OCR around 30 titles, so our digitization work took much less time and energy than we had anticipated.

¹⁴Anon., “75-plussers sterkst groeiende groep internetters,” in: *CBS* (December 27, 2016), <https://www.cbs.nl/nl-nl/nieuws/2016/52/75-plussers-sterkst-groeiende-groep-internetters> (retrieved on June 11, 2019).

Third, our driving force was always the research questions we have, and we did not intend to limit ourselves to the methods that had been used in previous literary studies. Our curiosity thus led us to cross boundaries. Interdisciplinary and transdisciplinary collaboration was therefore key in our project. First, we made sure that our team consisted of people with a mix of disciplinary backgrounds and expertise. The core team had a background in literary studies, computational linguistics, natural language processing, stylometry, lexicology and computer science. Second, we initiated collaborations with people from other disciplines, such as psychology and statistics, and from libraries and publishers, and with translators; sometimes, people approached us to initiate a joint project. And each of these collaborations led to new ideas. But collaboration only works well when the ultimate question is interesting for all those involved, so that each collaborator identifies with the shared research and research questions and has a stake in the outcome or some of the outcomes. From the first preparations for the project onwards, we found that we were not the only ones interested in the question of literary quality. This made collaboration a lot easier. For many of our collaborators there were also interesting results to expect, and we may need many more years to follow up on all the ideas and plans that were shaped.

Finally, this kind of collaboration needs to be on an equal footing, which implies that we need to share the results and ideas with each other. Perhaps this is the most difficult issue for many of us, as we come from a branch of humanities where the lonesome individual genius is still highly respected.

To sum up: I am convinced that the rise of information and communication technology has greatly enhanced the possibilities for humanities research—as long as researchers are willing to cross disciplinary boundaries themselves or are willing to collaborate on an equal footing (which also implies crossing at least some boundaries) with scholars from different disciplines. In the near future, these kinds of developments may greatly change the location of these disciplinary boundaries and even change the way that humanities research fields are organized.

References

- All digital references were last viewed on June 11, 2019.
- Anon., “75-plussers sterkst groeiende groep internetters,” in: *CBS* (December 27, 2016), <https://www.cbs.nl/nl-nl/nieuws/2016/52/75-plussers-sterkst-groeiende-groep-internetters>.
- Anthony, Laurence, *AntConc* (Version 3.4.3), Tokyo 2014, <http://www.laurenceanthony.net/>.
- Brems, Elke/Réthelyi, Orsolya/Kalmthout, Ton van (eds), *Doing Double Dutch: The international circulation of literature from the Low Countries*, Leuven 2017.
- Casanova, Pascale, *The world republic of letters*, translated by M.B. DeBevoise, Cambridge, MA/London 2004.
- Cranenburgh, Andreas van, *Rich statistical parsing and literary language*, Amsterdam 2016, <http://hdl.handle.net/11245/1.543163>.
- Dalen-Oskam, Karina van, “Prehistory of The Riddle” (“The Riddle of Literary Quality: The search for conventions of literariness,” transl. of: “The Riddle of Literary Quality. Op zoek naar conventies van literariteit,” in: *Vooys: tijdschrift voor letteren* 32/3 [2014], 25–33), <http://literaryquality.huygens.knaw.nl/?p=537#more-537>.

- Franssen, Thomas, *How books travel. Translation flows and practices of Dutch acquiring editors and New York literary scouts, 1980–2009*, Amsterdam 2015, <https://hdl.handle.net/11245/1.446961>.
- Koolen, Corina, *Reading beyond the female: The relationship between perception of author gender and literary quality*, Amsterdam 2018, <https://hdl.handle.net/11245.1/cb936704-8215-4f47-9013-0d43d37f1ce7>.
- Wilterdink, Nico, “Breaching the dyke. The international reception of contemporary Dutch translated literature,” in: Elke Brems/Orsolya Réthelyi/Ton van Kalmthout (eds), *Doing Double Dutch: The international circulation of literature from the Low Countries*, Leuven 2017, 45–65.
- Wouden, Ton van der/Lensink, Saskia/Dalen-Oskam, Karina van, “Hidden indicators. Modal Particles in modern Dutch fiction and fiction translated from English: a quantitative approach“ (forthcoming).

Online Resources

- The National Reader Survey*, <https://www.hetnationalelezersonderzoek.nl/>.
- The Riddle of Literary Quality*, <http://literaryquality.huygens.knaw.nl/>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Briefeditionen vernetzen

Stefan Dumont

1 Ausgangslage

Briefe¹ gehören zu den wertvollsten Quellen für die literatur-, geschichts-, musik- und sprachwissenschaftliche Forschung.² Im Gegensatz zu veröffentlichten Monographien, Aufsätzen, Partituren etc. versprechen Briefe einen intimeren oder unverfälschteren Einblick in die Gedankenwelt von Menschen. Denn Briefe gehören zu den Textsorten, die vom Autor meistens erst einmal nicht für eine Veröffentlichung vorgesehen waren.³ Das hat auch zur Folge, dass Briefe eine Ad-hoc-Schriftlichkeit abbilden und damit eine gewisse Nähe zur Mündlichkeit bekommen.⁴ Neben die (mehr oder weniger vermeintliche) Privatheit tritt ein weiterer Aspekt: Im Gegensatz zu Aufsätzen, Abhandlungen, Romanen etc., die über Titel und ggf. Inhaltsangaben verfügen, kann der Inhalt eines Briefes von der Wissenschaftlerin kaum antizipiert werden, bevor sie ihn gelesen hat. Hinzu kommt, dass der Inhalt häufig sehr heterogen ist: Im Brief werden ganz unterschiedliche Ereignisse, Erkenntnisse, Personen oder Publikationen aus der

¹Mit ‚Brief‘ sind in diesem Aufsatz neben dem (Privat-)Brief im eigentlichen Sinne auch alle anderen vergleichbaren Schriftstücktypen gemeint, wie z. B. Postkarten, Billets und Telegramme sowie Schreiben amtlichen Charakters.

²Reinhard M. G. Nickisch, *Brief*, Stuttgart 1991, 59.

³Gert Mattenklott/Hannelore Schlaffer/Heinz Schlaffer (Hg.), *Deutsche Briefe 1750–1950*, Frankfurt a. M. 1988, 7.

⁴Stephan Elspaß, *Sprachgeschichte von unten. Untersuchungen zum geschriebenen Alltagsdeutsch im 19. Jahrhundert*, Berlin/Boston 2011, 24–27, <https://doi.org/10.1515/9783110910568>.

S. Dumont (✉)
Berlin-Brandenburgische Akademie der Wissenschaften, Berlin, Deutschland
E-Mail: dumont@bbaw.de

Lebens- und Arbeitswelt der Schreiber angesprochen und kommentiert. Irmtraud Schmid spricht gar von einem „Gemischtwarenladen“.⁵ Gerade aus diesen Gründen ist die moderne Wissenschaft an Briefen als Quellen äußerst interessiert. So untersucht z. B. die Wissenschaftsgeschichte anhand von Briefen, wie Erkenntnisse gewonnen, diskutiert und verbreitet wurden. Die Musikwissenschaft geht anhand von Briefen der Entstehung von Kompositionen nach, während die Politikgeschichte wiederum darin nach Ansichten und Bewertungen von politischen Ereignissen und Theorien fahndet. Die Literaturwissenschaft schließlich möchte Werkentstehungen untersuchen oder behandelt den Brief selbst als literarisches Erzeugnis. Darüber hinaus interessieren sich Wissenschaftlerinnen für Briefe, weil ein Brief „so gut wie nie als nomadisches Einzelobjekt existiert“,⁶ sondern einerseits in einer Reihe von vorhergehenden oder nachfolgenden Briefen steht, andererseits die Korrespondenten fast immer nicht nur mit einem, sondern mit mehreren Partnern im epistolaren Austausch stehen. Damit bilden mehrere Briefwechsel ein Briefnetz – ein Aspekt, der die Textsorte ‚Brief‘ von den meisten anderen abhebt und auch einen eigenen Zugriff nötig macht. Der Erforschung dieser Briefnetze widmet sich heute – neben den bereits genannten Disziplinen – insbesondere auch die historische Netzwerkforschung.

Das hohe wissenschaftliche Interesse führte dazu, dass Briefe seit Langem für die Forschung erschlossen werden. Zum einen geschieht das über archivalische Nachweismittel. Neben den üblichen Findbüchern und Datenbanken der einzelnen Archive ist hier insbesondere die *Zentralkartei der Autographen (ZKA)* zu nennen, die von 1966 bis zur Jahrtausendwende von der *Staatsbibliothek zu Berlin* gepflegt wurde. 2001 ist die ZKA im elektronischen Katalog *Kalliope* aufgegangen und wird dort weiterhin gepflegt, nun aber in einer Verbundarchitektur. *Kalliope* weist derzeit neben anderen Dokumenttypen mehr als zwei Millionen Briefe in (vornehmlich deutschen) Archiven und Bibliotheken nach.⁷ Die Erfassung von Briefen erfolgt dezentral in den Einrichtungen, die Erschließungstiefe variiert dabei von den reinen Basisdaten eines Briefes (Absender, Empfänger, Schreibdatum) über eine grobe Verschlagwortung bis hin zu einem kurzen Regest. Im Großteil der Fälle bleibt es allerdings bei den Basisdaten, da eine tiefere Erschließung für die Archive und Bibliotheken zu zeitaufwendig wäre. Fragen nach Briefinhalten, Datierungen oder Identifizierung von Absendern können in diesem Rahmen nicht geklärt werden.

⁵ Irmtraud Schmid, „Anforderungen an die Kommentierung von Briefen und amtlichen Schriftstücken“, in: *„Ich an Dich“*. Edition, Rezeption und Kommentierung von Briefen, Innsbruck 2001, 35–45, hier: 38.

⁶ Wolfgang Bunzel, „Briefe, Briefnetze, Briefnetzwerke. Überlegungen zur epistolaren Interkonnektivität“, in: *Fontanes Briefe ediert*, Würzburg 2014, 232–245, hier: 234.

⁷ Jutta Weber, „Briefnachlässe auf dem Wege zur elektronischen Publikation. Stationen neuer Beziehungen“, in: Anne Bohnenkamp/Elke Richter (Hg.), *Brief-Edition im digitalen Zeitalter*, Berlin/Boston 2013, 25–34.

Zum anderen werden Briefe aber seit dem 19. Jahrhundert in gedruckten Briefausgaben für die weitere Forschung bereitgestellt. Die Bandbreite reicht von reinen Regestaussagen, die lediglich Zusammenfassungen der einzelnen Briefe enthalten, über mehr oder weniger genaue Transkriptionen bis hin zur vollständigen Edition mit textkritischem Kommentar und inhaltlichen Erläuterungen. Die *Deutsche Forschungsgemeinschaft (DFG)* fördert seit Langem die Erstellung solcher Editionen und hat dazu schon früh einige Kolloquien abgehalten. Zu den zahlreichen Gesamtausgaben, die heute vor allem im Rahmen des Akademienprogramms erstellt werden, gehören Briefeditionen selbstverständlich dazu.

Diese Masse an gedruckten Briefen veranlasste bereits 1969 Fritz Schlawe dazu, eine *Bibliographie der Briefausgaben und Gesamtregister der Briefschreiber und Briefempfänger der Briefsammlungen des 19. Jahrhunderts* vorzulegen. Seine Begründung: „Von den Briefen weniger großer Persönlichkeiten abgesehen, deren *œuvre* in Gesamtausgaben vorliegt oder durch Spezialbibliographien erschlossen ist, waren diese Briefe bis jetzt zum großen Teil bibliographisch nicht näher nachweisbar und damit weithin der Auswertung entzogen“.⁸ In seiner Bibliographie erfasste er rund 20.000 Korrespondenzen. Schon damals ob der ungeheuren Menge unvollständig, ist diese Bibliographie nach fast 45 Jahren stark veraltet.

Gedruckte Briefeditionen waren (und sind) so angelegt, dass fast immer nur die Briefe von bzw. an eine Person oder gar nur der Briefwechsel zwischen zwei bestimmten Personen aufgenommen wird. In neueren, genuin digital angelegten Briefeditionen werden gelegentlich darüber hinaus auch Briefe berücksichtigt, die im Zentrum des Editionsvorhabens stehende Persönlichkeit ‚nur‘ zum Inhalt haben, aber nicht deren Korrespondenz entstammen. Außerdem gibt es auch digitale Editionen, die Korrespondenzen thematisch orientiert zu erschließen versuchen, so z. B. das Projekt *Briefe und Texte aus dem intellektuellen Berlin um 1800*.⁹ Die große Masse an Editionen geht aber nicht so vor. Insbesondere die Gesamtausgaben stehen vor dem Problem, dass die Fokussierung auf eine Person deren Bedeutung bzw. die ihrer Korrespondenzpartner verfälschen kann.¹⁰ Demgegenüber hat die Forschung in den letzten Jahrzehnten verstärkt das *Umfeld* der untersuchten Persönlichkeiten in den Blick genommen.

Gegenüber archivalischen Findmitteln und thematisch orientierten Nachweisdatenbanken bieten historisch-kritische Briefeditionen die beste Erschließungstiefe, weil die Texte durch Transkription, Kommentierung und Registrierung systematisch und vollständig bearbeitet und erschlossen wurden. Doch dieses Wissen steht nur dann der Wissenschaftlerin zur Verfügung, wenn

⁸Fritz Schlawe, *Die Briefsammlungen des 19. Jahrhunderts. Bibliographie d. Briefausg. u. Gesamtregister d. Briefschreiber u. Briefempfänger; 1815–1915*, Stuttgart 1969, XII.

⁹Anne Baillot (Hg.), *Berliner Intellektuelle. Briefe und Texte aus dem intellektuellen Berlin um 1800*, Berlin, <http://www.berliner-intellektuelle.eu/> (letzter Aufruf 24.09.2018).

¹⁰Heinz Becker, „Die Briefausgabe als Dokumenten-Biographie“, in: Wolfgang Frühwald/Hans-Joachim Mähl/Walter Müller-Seidel (Hg.), *Probleme der Brief-Edition. Kolloquium der Deutschen Forschungsgemeinschaft (DFG)*, Boppard 1977, 11–25, hier: 16.

ihr die Edition eines einschlägigen Briefes bekannt ist. Wenn dies nicht der Fall ist, muss die Wissenschaftlerin recherchieren – und das ist zeitaufwendig, insbesondere wenn dem gesuchten Korrespondenten nicht eine eigene Edition gewidmet wurde. Will man darüber hinaus aber Briefinhalte finden, ist eine arbeitsökonomische Recherche fast schon unmöglich. Dabei benutzen Wissenschaftlerinnen Briefeditionen tatsächlich häufig sehr selektiv, da sie an ganz bestimmten Details oder Fragen interessiert sind. So konstatiert auch Jochen Strobel, „[...] dass Briefleser oder besser Nutzer von Briefeditionen nur in Ausnahmefällen gewillt sind, dicke Bände zu lesen, in denen sich unüberschaubar heterogene Daten verbergen, die meisten davon für das spezifische Nutzerinteresse noch dazu bedeutungslos“.¹¹ Und Norbert Oellers fragte schon 1993 rhetorisch: „Aber wer liest einen Briefband wie einen Roman?“¹²

Und so kommt Wolfgang Bunzel – mit Blick auf die Erforschung der Romantik – zum Schluss:

Die schiere Masse der überlieferten Briefe ist selbst für Experten nicht zu überschauen, viel weniger für Forscher, die sich nur mit einzelnen Personen beschäftigen oder an Detailspekten interessiert sind. Dies hat zur Folge, dass schon das gedruckt vorliegende Material nur höchst selektiv benutzt wird und die Mehrzahl der Briefe unzitiert, meist sogar ungesichtet bleibt.¹³

Spätestens seit der Jahrtausendwende wird in Bezug auf Briefe über eine Lösung dieses Problems mithilfe der Informationstechnologie diskutiert und auch tatsächlich verfolgt. So initiierte 2005 das Heidelberger Institut für Textkritik den *BRIEFkasten* – eine Datenbank, die wie Schlawes Publikation gedruckte bzw. edierte Briefe nachweisen wollte.¹⁴ Weitere IT-gestützte Projekte wurden vor allem zur Erforschung der *Res publica literaria* zwischen 1550 und 1750 initiiert. Zu nennen sind hier insbesondere *Early Modern Letters Online (EMLO)*¹⁵ sowie *Frühneuzeitliche Ärztebriefe des deutschsprachigen Raums*¹⁶. Diese beiden Projekte haben gemein, dass sie Datenbanken erstellen, die Gelehrtenbriefe – ob archivalisch oder gedruckt überliefert oder nur erschlossen – aus der Frühen Neuzeit nachweisen; neben der entsprechenden Verschlagwortung enthalten

¹¹ Jochen Strobel, „Der Brief als Prozess. Entwurf und Konzept in der digitalen Edition“, in: Anne Bohnenkamp (Hg.), *Brief-Edition im digitalen Zeitalter*, Berlin/Boston 2013, 133–146, hier: 136.

¹² Norbert Oellers, „Wie sollten Briefwechsel ediert werden?“, in: Lothar Bluhm/Andreas Meier (Hg.), *Der Brief in Klassik und Romantik*, Würzburg 1993, 1–12, hier: 8.

¹³ Wolfgang Bunzel, „Briefnetzwerke der Romantik. Theorie – Praxis – Edition“, in: Anne Bohnenkamp/Elke Richter (Hg.), *Brief-Edition im digitalen Zeitalter*, Berlin/Boston 2013, 109–132, hier: 117.

¹⁴ *BRIEFkasten*, <http://www.textkritik.de/briefkasten/bkprojekt.htm> (letzter Aufruf 24.09.2018).

¹⁵ <http://emlo.bodleian.ox.ac.uk/> (letzter Aufruf 24.09.2018).

¹⁶ <http://www.medizingeschichte.uni-wuerzburg.de/akademie/index.html> (letzter Aufruf 24.09.2018).

sie teilweise auch Regesten der aufgenommenen Briefe. Darüber hinaus versuchen gerade die Projekte zur Erforschung der *Res publica literaria*, den schon angesprochenen Netzcharakter der Korrespondenzen für die Forschung nutzbar zu machen.¹⁷ Während *EMLO* teilweise kollaborativ erweitert wird, entsteht die Datenbank *Frühneuzeitliche Ärztebriefe* im Rahmen eines Akademienvorhabens an der *Bayerischen Akademie der Wissenschaften*. Die in beiden Projekten gesammelten Briefmetadaten stehen zur kostenlosen Recherche bereit, können aber nicht über technische Schnittstellen unter einer freien Lizenz bezogen und nachgenutzt werden.¹⁸

Für andere Forschungsthemen und Epochen, wie etwa die deutsche Romantik oder das 19. Jahrhundert insgesamt, stehen entsprechende Datenbankprojekte aber nicht zur Verfügung. Das ist für die Forschung eine missliche Lage, erreicht das Zeitalter des deutschsprachigen Privatbriefs doch in der ersten Hälfte des 19. Jahrhunderts erst seinen Höhepunkt.¹⁹ Angesichts der enormen Briefmengen, die zahlreiche Editionsprojekte herausgeben, scheint aber eine ‚klassische‘, zentral verwaltete Datenbank aller edierten Briefe schwierig zu realisieren, insbesondere vor dem Hintergrund, dass die Anzahl edierter Briefe durch Neuvorhaben stetig anwächst.²⁰

Und so fordert auch Bunzel:

die Schaffung einer dezentralen, möglichst offenen, auf [...] html/xml-Grundlage basierenden und mit TEI-Minimalstandards operierenden digitalen Plattform, die nach vielen Richtungen hin erweiterbar ist und es den bereits bestehenden Portalen und Homepages erlaubt, sich mit denkbar geringem Zusatzaufwand daran zu beteiligen. Nötig ist dafür keine Superstruktur, welche die – ohnehin nicht exakt bezifferbare – Gesamtheit aller Briefe der Romantik überwölbt, sondern vielmehr ein intelligentes Verknüpfungssystem, das vorhandene Dokumente in Konnex zueinander bringt.²¹

Die nötigen Grundvoraussetzungen für ein solches Verknüpfungssystem hatte Andrea Rapp bereits 2007 auf einer Tagung des DFG-Projekts *Entwicklung von Werkzeugen für digitale Formen wissenschaftlich-kritischer Musikeditionen* genannt:

¹⁷Die weitergehende Erforschung der *Res publica literaria* ist auch Gegenstand des EU COST-Projekts *Reassembling the Republic of Letters*, <http://www.republicofletters.net/> (letzter Aufruf 24.09.2018).

¹⁸In der gerade begonnenen neuen Förderphase von *EMLO* sind anscheinend Schnittstellen bzw. Webservices für Daten zu Personen, Orten und Datumsangaben geplant, aber – soweit ersichtlich – nicht für die Briefmetadaten. Vgl. <http://www.culturesofknowledge.org/?p=8455> (letzter Aufruf 24.09.2017).

¹⁹Nickisch (Anm. 2), 54.

²⁰Jürgen Herres/Manfred Neuhaus, „Vorwort der Herausgeber“, in: Jürgen Herres/Manfred Neuhaus (Hg.), *Politische Netzwerke durch Briefkommunikation. Briefkultur der politischen Oppositionsbewegungen und frühen Arbeiterbewegungen im 19. Jahrhundert*, Berlin 2002, 7–25, hier: 18.

²¹Bunzel (Anm. 13), 123. Vgl. auch Øyvind Eide, „Putting the Dialogue Back Together Re-Creating Structure in Letter Publishing“, in: *Computers and the Humanities* 37/1 (2003), 65–75, hier: 74. Eide fordert ebenfalls die Vernetzung von Briefeditionen untereinander.

Standardisierung ist eine wichtige Voraussetzung, da sich erst mit ihr Perspektiven der Vernetzung auf tun; Die Text Encoding Initiative (TEI) bietet hier einen flexiblen Standard, um für das eigene Corpus wichtige Details zu codieren. Darüber hinaus kann für die Vernetzung – und speziell für Suchanfragen – aber auch ein baseline-encoding sinnvoll sein, [...] d. h. die Einigung auf einen kleinsten gemeinsamen Nenner, der für Briefe z. B. in der Angabe von Absender, Empfänger, Ort und Datum bestehen könnte. Schon diese wenigen Informationen erleichtern die Suche über große Corpora hinweg. Denkbar wäre dadurch etwa die Schaffung eines projekt- und länderübergreifenden Briefverzeichnisses, dem dann später die Brieftexte angefügt werden könnten. Damit eröffnen sich z. B. Perspektiven für die Untersuchung von Korrespondentenbeziehungen mit neuartigen Fragestellungen.²²

Die Frage nach solchen gemeinsamen Nennern und einem übergreifenden Vernetzungssystem für Briefeditionen wurde 2014 auch in einem Workshop von Anne Baillet und Markus Schnöpf zum Thema *Briefeditionen um 1800. Schnittstellen finden und vernetzen* aufgegriffen. Auf diesem Workshop entstand die Initiative zur Entwicklung des *Correspondence Metadata Interchange Format* und zum Webservice *correspSearch*. Die derzeit online verfügbare Basisversion des Webservices²³ wurde von der DH-Arbeitsgruppe *TELOTA* an der *Berlin-Brandenburgischen Akademie der Wissenschaften* entwickelt. Seit Herbst 2017 wird *correspSearch* nun in einem von der DFG geförderten Projekt weiterentwickelt (Abb. 1).

2 Austauschformat für Briefmetadaten

Grundprinzip des Webservice *correspSearch* ist es, die Briefmetadaten aus digitalen Briefverzeichnissen, die von den Editionsprojekten selbst bereitgestellt werden, zu beziehen und zu aggregieren. Es wurde also keine Datenbank im klassischen Sinne geschaffen, in der die Daten nur einmalig aus ggf. unterschiedlichen Quellformaten importiert werden und dann nur in einer zentralen Datenbank weitergepflegt werden können. Stattdessen wurde – wie von Bunzel vorgeschlagen – mit *correspSearch* ein offenes System entwickelt, das lediglich bereits vorhandene Daten, die in einem standardisierten Austauschformat auf TEI-XML-Basis vorliegen, aggregiert. Letzteres ist das *Correspondence Metadata Interchange Format (CMIF)*, das von der *TEI Correspondence Special Interest Group* entwickelt wurde und wird.²⁴

²²Andrea Rapp, „Einige Anmerkungen zu Retrodigitalisierungs-Verfahren und Perspektiven digitaler Briefeditionen“, in: Peter Stadler/Joachim Veit (Hg.), *Digitale Edition zwischen Experiment und Standardisierung. Musik – Text – Codierung*, Tübingen 2009, 203–206, hier: 205 f. Auf die Notwendigkeit von standardisierten Metadaten zur besseren Auffindbarkeit von edierten Briefen verweist auch Jochen Strobel, „Die Normierung von Metadaten als Standardisierungsinstrument in der digitalen Briefedition“, in: Michael Stolz/Yen-Chun Chen (Hg.), *Internationalität und Interdisziplinarität der Editionswissenschaft*, Berlin/Boston 2014, 99–107, hier: 104 f.

²³<https://correspSearch.net> (letzter Aufruf 24.9.2017).

²⁴TEI Correspondence SIG (Hg.), *Correspondence Metadata Interchange Format (CMIF) 2015*, <https://github.com/TEI-Correspondence-SIG/CMIF> (letzter Aufruf 16.09.2017).

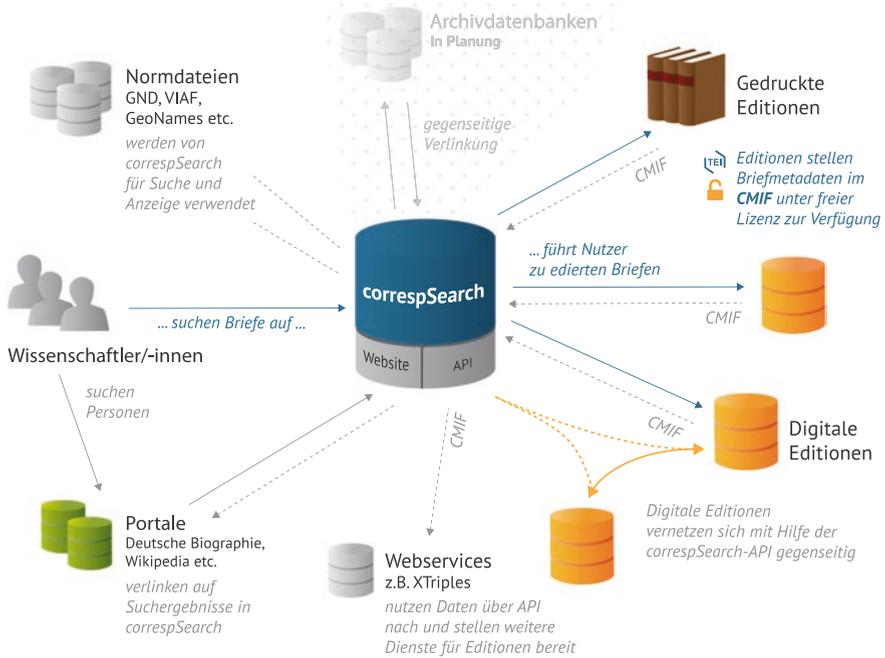


Abb. 1 Überblick über den Webservice *correspSearch*

Im Kern basiert das CMIF auf dem TEI-Elementset <correspDesc> (Correspondence Description), das von einer Taskforce – namentlich Marcel Illetschko, Sabine Seifert und Peter Stadler – initiiert und von 2013 bis 2015 entwickelt wurde. Dabei dient <correspDesc> ausschließlich dazu, die kommunikationsspezifischen Metadaten eines Briefes zu erfassen, d. h. insbesondere Absender, Empfänger, Schreib- und Empfangsdatum bzw. -ort. Darüber hinaus kann der Korrespondenzkontext (auch ‚Korrespondenzstelle‘ genannt) innerhalb eines Briefwechsels angegeben werden – also der Brief, auf den geantwortet wird, bzw. der Brief, mit dem wiederum der vorliegende beantwortet wird.²⁵ Alle anderen Metadaten, die typischerweise beim Edieren eines Briefes anfallen, werden mit den dafür schon länger vorhandenen TEI-Elementen (msDesc, physDesc etc.) kodiert. Das TEI-Elementset <correspDesc> umfasst

²⁵ Peter Stadler/Marcel Illetschko/Sabine Seifert, „Towards a Model for Encoding Correspondence in the TEI. Developing and Implementing <correspDesc>“, in: *Journal of the Text Encoding Initiative [Online]* 9 (2016), <https://doi.org/10.4000/jtei.1433>.

also, grob gesagt, diejenigen Daten des Briefes, die in Briefeditionen i. d. R. auch als ‚Kopfdaten‘ oder ‚Titel‘ des Briefes notiert werden. Es sind auch gleichzeitig jene Daten, die in Briefeditionen im Briefverzeichnis aufgelistet werden. Deshalb lag es nahe, das Elementset <correspDesc> nicht nur zur Erfassung der Metadaten eines einzelnen Briefes zu verwenden, sondern auch im Rahmen eines digitalen Briefverzeichnisses.²⁶

Im Unterschied zur Spezifikation in den TEI-Richtlinien wird <correspDesc> im CMIF allerdings deutlich restriktiver und in reduzierter Form verwendet. So ist es z. B. nach den TEI-Richtlinien durchaus möglich, eine komplette Adresse mitsamt detaillierter Auszeichnung im Kindelement <correspAction> zu notieren. Im CMIF ist das nicht erlaubt, vielmehr beschränkt sich die Angabe hier auf die Elemente <persName>, <orgName> und <placeName>, um die Briefpartner und ihre Aufenthaltsorte festzuhalten. Eine Unterstrukturierung der Elemente ist nicht erlaubt; dafür wird im CMIF empfohlen, zusätzlich eine ID aus einer Normdatei zu verwenden, um Personen und Orte zu identifizieren (dazu unten mehr). Auch im Fall der Datumsangaben ist die Verwendung gegenüber den TEI-Richtlinien eingeschränkt: Zugelassen sind (derzeit) nur die Attribute @when, @from, @to, @notBefore, @notAfter. Darüber hinaus können nicht alle vom W3C vorgesehenen Datumswerte notiert werden, sondern nur die gängigsten.

Das Austauschformat reduziert also die Ausdrucksmöglichkeiten und schränkt damit Umfang und Genauigkeit der Daten zwangsläufig ein. Auf der einen Seite bedeutet dies natürlich einen Nachteil, weil nicht alle in der Edition vorhandenen Metadaten auf diesem Wege bereitgestellt werden können. Die Reduktion und Restriktion ist aber notwendig, um die Metadaten durch einen Webservice wie *correspSearch* verarbeiten zu können. Zum einen sollte nämlich die Anzahl der Metadatenfelder, die in der Suche ausgewertet werden müssen, überschaubar bleiben und sich an den gängigsten Anforderungen orientieren. Zum anderen muss die Kodierung eindeutig sein. Die Verwendung von TEI-XML allein gewährleistet dies nämlich noch nicht. Zwar harmonisiert sie schon in gewissem Maße digitale Editionen, allerdings handelt es sich bei der TEI um „Richtlinien“ („Guidelines“) und nicht um einen ‚Standard‘ im engeren und eigentlichen Sinne. Letzterer müsste dafür sorgen, dass gleichartige Textphänomene auch gleichartig und eindeutig kodiert werden. Dies ist aber nicht der Fall, weshalb es auch Unterformate, wie z. B. das *Basisformat des Deutschen Textarchivs (DTABf)* gibt, die eine entsprechende Eindeutigkeit der Kodierung herstellen.²⁷

Im Bereich der Briefeditionen stellt sich die Ausgangssituation vergleichsweise gut dar: bereits seit drei Jahrzehnten wird in den Editionswissenschaften

²⁶Peter Stadler, „Interoperabilität von digitalen Briefeditionen“, in: Hanna Delf von Wolzhagen (Hg.), *Fontanes Briefe ediert*, Würzburg 2014, 278–287, hier: 284–286.

²⁷Susanne Haaf/Alexander Geyken/Frank Wiegand, „The DTA ‚Base Format‘“. A TEI Subset for the Compilation of a Large Reference Corpus of Printed Text from Multiple Sources“, in: *Journal of the Text Encoding Initiative* 8 (2015), <https://doi.org/10.4000/jtei.1114>.

über Standardisierung, Vereinheitlichung und Vernetzung von Briefeditionen diskutiert.²⁸ Diese Diskussion zeigt zum einen den Bedarf nach Standardisierung, zum anderen lässt sich gut auf dieser Basis aufbauen. Darüber hinaus hat die Verwendung eines Austauschformats für Briefmetadaten einen entscheidenden Vorteil: Das digitale Briefverzeichnis wird *zusätzlich*, d. h. *separat* zur Edition, bereitgestellt. Man muss eben nicht die Edition (d. h. die Kodierung der edierten Texte) an sich standardisieren, sondern nur die Schnittstellen zwischen den Editionen.²⁹ So kann in der einzelnen digitalen Briefedition eine projektspezifische Kodierung gewählt werden, die dem zu edierenden Material und dem Editionsprojekt gerecht wird. Das Editionsprojekt muss lediglich darauf achten, dass mit den vorhandenen Daten ein digitales Briefverzeichnis als CMIF bereitgestellt werden kann.

Die Konzeption, die Briefmetadaten einer Edition (oder auch einer Briefdatenbank) über eine separate Datei in einem Austauschformat bereitzustellen, hat den Vorteil, dass die Metadaten prinzipiell stets in der Obhut des Projekts oder der Projektträgerin verbleiben. Der Webservice bezieht in regelmäßigen Abständen diese CMIF-Dateien neu und aktualisiert so den abfragbaren Bestand an Metadaten. Dadurch können Briefmetadaten bei Bedarf (z. B. im Fall neuer Erkenntnisse) einfach vom jeweiligen Editionsprojekt ergänzt, aktualisiert und korrigiert werden. Außerdem müssen – im Fall digitaler Briefeditionen – die Daten auch nicht zusätzlich zur Edition in einem anderen System gepflegt werden: die CMIF-Datei wird einfach aus der digitalen Edition heraus über ein Skript erstellt.

Das Austauschformat ist öffentlich spezifiziert und umfassend dokumentiert, sodass es jedem Editionsprojekt leicht möglich ist, eigene Daten bereitzustellen. Dafür muss die CMIF-Datei lediglich online erreichbar und unter einer freien Lizenz gestellt sein. Die freie Lizenzierung ermöglicht eine rechtssichere Weiterverwendung der Daten nicht nur im Webservice *correspSearch*, sondern auch in anderen Datenbanken, Projekten und digitalen Briefeditionen.

Das CMI-Format wird von der *TEI Correspondence SIG* in Zukunft weiterentwickelt und um weitere Kodierungsmöglichkeiten ergänzt werden, u. a. um erwähnte Entitäten (siehe nächster Abschnitt) oder die zugrundeliegende Archivalie notieren zu können.³⁰

²⁸ Winfried Woesler, „Vorschläge für eine Normierung von Briefeditionen“, in: *editio* 2 (1988), 8–18; Bernhard Appel/Werner Breig/Gabriele Buschmeier u. a. (Hg.), *Richtlinien-Empfehlungen zur Edition von Musikerbriefen. Im Auftrag der Arbeitsgemeinschaft Musikerbriefe innerhalb der Fachgruppe Freie Forschungsinstitute in der Gesellschaft für Musikforschung*, Mainz 1997; Strobel (Anm. 21).

²⁹ Stadler (Anm. 22), 285.

³⁰ Einen ersten Überblick über die mögliche Weiterentwicklung gibt Stefan Dumont, „Perspectives of the further development of the Correspondence Metadata Interchange Format (CMIF)“, in: *digiversity – Webmagazin für Informationstechnologie in den Geisteswissenschaften* (2015), <https://digiversity.net/2015/perspectives-of-the-further-development-of-the-correspondence-metadata-interchange-format-cmif/> (letzter Aufruf 16.9.2017).

3 Identifizierung und Adressierung

Ein standardisiertes Austauschformat ist nur eine wichtige Voraussetzung, um Briefeditionen anhand eines Webservices wie *correspSearch* zu vernetzen. Die zweite notwendige Bedingung ist die technische Möglichkeit, *projektübergreifend* Entitäten zu identifizieren und zu adressieren. Entitäten können konkret oder abstrakt, materiell oder immateriell sein. Dazu gehören z. B. Personen, Körperschaften, Orte, Ereignisse, Werke oder Gegenstände. Die Identifizierung von Entitäten ist in Briefeditionen wichtig – zum einen weil Personen und Orte als Teilnehmer des Kommunikationsereignisses ‚Brief‘ auftreten, zum anderen weil in Briefen, wie eingangs skizziert, häufig unterschiedliche Personen, Werke etc. vom Briefschreiber erwähnt und kommentiert werden –, ohne dass dies auf den ersten Blick ersichtlich ist. Gerade Briefeditionen werden von Wissenschaftlerinnen selektiv nach diesen Erwähnungen durchsucht und konsultiert. Daher wurden auch schon vor dem digitalen Zeitalter die Register als unverzichtbarer Bestandteil von Briefeditionen gesehen.³¹

Allerdings hat sich das im Druckzeitalter verwendete System von Namen (ggf. in Kombination mit weiteren Angaben, wie z. B. bei Personen deren Lebensdaten), Titeln und Signaturen für die maschinelle Verarbeitung als fehleranfällig und ungünstig erwiesen, da diese Angaben entweder mehrdeutig oder unterschiedlich notiert sein können. Deshalb wurden und werden in Datenbanken numerische oder alphanumerische Kennungen für Entitäten vergeben. Da diese Kennungen immer nur in einem Bezugssystem (z. B. einer Datenbank oder einem Projekt) gültig sind, ist es notwendig, auch die Instanz anzugeben, die die konkret vorliegenden Kennungen verwaltet und in deren Rahmen die Kennungen gelten.

Dieses Prinzip liegt auch dem Konzept des *Uniform Resource Identifiers (URI)* zugrunde, das Tim Berners-Lee 1994 einführte und das durch das *W3C* zu einem Standard weiterentwickelt wurde.³² Neben einer festgelegten Syntax, die die Maschinenlesbarkeit garantiert, basiert der Grundgedanke von URIs auch auf der Forderung nach Permanenz, also dauerhafter Verfügbarkeit bzw. Adressierbarkeit der URI.³³ Beide Vorteile machen URIs unverzichtbar für digitale Daten und Publikationen.

Daher gehören URIs heute selbstverständlich zum Web und werden vielfach angeboten und verwendet. So bieten Archive, Bibliotheken oder Verbundkataloge (z. B. *Kalliope*) permanente URIs an, um archivalische Materialien, Bücher und

³¹ Uta Motschmann, „Von den Registern“, in: Siegfried Scheibe/Waltraud Hagen/Christel Laufer u. a. (Hg.), *Vom Umgang mit Editionen. Eine Einführung in Verfahrensweisen und Methoden der Textologie*, Berlin 1988, 225–263, hier: 225 f.

³² Tim Berners-Lee/Roy Fielding/Larry Masinter, *RFC 3986. Uniform Resource Identifier (URI). Generic Syntax* 2005, <https://tools.ietf.org/html/rfc3986> (letzter Aufruf 24.9.2017).

³³ Tim Berners-Lee, *Cool URIs Don't Change* 1998, <https://www.w3.org/Provider/Style/URI> (letzter Aufruf 24.9.2017).

davon abgeleitete Digitalisate eindeutig und dauerhaft zu identifizieren. Die *Deutsche Nationalbibliothek (DNB)* bietet für jedes nach 1913 in Deutschland erschienene Buch eine solche URI an. Und auch Wissenschaftseinrichtungen beginnen zunehmend damit, digitale Publikationen (oder Teile davon) mit URIs zu versehen. Daher setzt auch *correspSearch* durchgehend auf die Verwendung von URIs. Generell empfiehlt es sich auch für digitale Briefeditionen, „cool URIs“ im Sinne von Berners-Lee bereitzustellen.

Wie schon erwähnt, ist es im Bereich der Briefeditionen aber nicht nur wichtig, die Archivalien und die gedruckt bzw. digital publizierten Briefe zu adressieren, sondern vor allem auch die im Zusammenhang mit ihnen stehenden Entitäten. Für Personen und Körperschaften hat sich seit Längerem im Bereich der deutschsprachigen *Digital Humanities* (und insbesondere auch der digitalen [Brief-] Editionen) die Verwendung von Kennungen aus der *Gemeinsamen Normdatei (GND)* etabliert.³⁴ Die GND wird von der *Deutschen Nationalbibliothek*, allen deutschsprachigen Bibliotheksverbänden mit den angeschlossenen Bibliotheken sowie zahlreichen weiteren Einrichtungen kooperativ geführt. Mit ihrer Hilfe und dem BEACON-Austauschformat³⁵ ist es beispielsweise möglich, Informationen zu Personen in Webangeboten wie z. B. digitale Nachschlagewerke und Editionen automatisiert zu verknüpfen. Dafür muss lediglich die ID aus der GND in der eigenen Datenbank notiert und eine BEACON-Schnittstelle angeboten werden.

Aufgrund der bisher schon sehr erfolgreichen Vernetzung mit Hilfe von GND-IDs und BEACON-Schnittstellen werden in CMIF-Dateien bzw. im Webservice *correspSearch* ebenfalls GND-URIs zur Identifizierung von Personen und Körperschaften verwendet. Über die Gemeinsame Normdatei hinaus werden aber auch IDs aus anderen einschlägigen Normdateien verwendet. Im Moment sind dies die *Autorités der Bibliothèque Nationale de France (BNF)*, die Normdatei des *Library of Congress (LC)* und die Kennungen der *National Diet Library (NDL)* Japans. Mit Hilfe der *Virtual International Authority File (VIAF)* ist es dem Webservice *correspSearch* möglich, eine ID auf ihr entsprechendes Pendant in einer der anderen unterstützten Normdateien abzubilden. So kann ein Editionsvorhaben Kennungen aus der präferierten Normdatei verwenden – was insbesondere im internationalen Kontext nützlich ist.

Dass die Verwendung von IDs aus Normdateien die automatisierte, projektübergreifende Identifizierung von Absendern und Empfängern erfolgreich ermöglicht, konnte beim bisherigen Betrieb von *correspSearch* gezeigt werden. Allerdings ist sie auch nicht unproblematisch. Zum einen gibt es Dubletten in der GND, d. h. für eine Person liegen ggf. mehrere Datensätze (also IDs) vor. Zum anderen gibt es nicht für alle Personen oder Körperschaften einen entsprechenden Datensatz und

³⁴ Peter Stadler, „Normdateien in der Edition“, in: *editio* 26 (2012), 174–183.

³⁵ Jakob Voß/Matthias Schindler, *BEACON Link Dump Format 2014*, <http://gbv.github.io/beaconspec/beacon.html> (letzter Aufruf 16.6.2017).

damit eine Norm-ID, die verwendet werden kann. Gerade im Bereich von Briefeditionen kann es mitunter viele Personen geben, die aus verschiedenen Gründen nicht in der GND auftauchen. Im digitalen Briefverzeichnis³⁶ der *Weber-Gesamtausgabe* ist z. B. für etwa 30 % der Personen keine GND-URI vorhanden. Dieser Umstand muss allerdings nicht gleich zu Problemen bei der Vernetzung führen, weil darunter auch viele Personen sind, die tatsächlich nur in einer Edition auftauchen und die daher auch nicht projektübergreifend identifiziert werden müssen. Dies können z. B. Familienangehörige sein oder lediglich lokal in Erscheinung tretende Personen. Hier könnten in Zukunft in *correspSearch* die ‚lokalen‘ URIs der jeweiligen Edition verwendet werden.

Neue Datensätze können in der GND zwar prinzipiell angelegt werden, aber dies geschieht, da die GND kooperativ geführt wird, bisher i. d. R. mithilfe einer regionalen Bibliothek, z. B. einer Universitätsbibliothek. Dieser Workflow wird in vielen Fällen auch so realisiert, steht aber aus unterschiedlichsten (meist administrativen) Gründen nicht jedem Projekt offen. Hier wären bessere Ergänzungsmöglichkeiten und neue kooperative Redaktionsmöglichkeiten nicht nur wichtig, sondern sogar unabdingbar. Dieses Problem wurde seitens der *DNB* erkannt; sie arbeitet daher derzeit an einem Plan zur Weiterentwicklung der GND, der insbesondere die oben genannten Probleme adressiert und bessere Kooperationsmöglichkeiten für nicht-bibliothekarische Kultur- und Wissenschaftseinrichtungen schaffen möchte.³⁷

Neben Personen und Orten stellt die *Deutsche Nationalbibliothek* in der GND auch Körperschaften und – in begrenztem Umfang – auch Werktitel zur Verfügung. Als Nationalbibliothek stellt sie natürlich auch URIs für Publikationen zur Verfügung; dies geschieht aber nicht in einer Normdatei (der GND), sondern über den Katalog der *DNB*. Dort sind alle seit 1913 in Deutschland erschienenen Publikationen verzeichnet. Für die Zeit vor 1913 ist die *DNB* nicht zuständig und führt daher auch keine Datensätze für Publikationen, die vor diesem Datum erschienen sind. Die Sammlung, Erschließung und der Nachweis älterer deutschsprachiger Bücher und Schriften geschieht kooperativ durch mehrere Bibliotheken. Die Nachweise werden als *Verzeichnisse der im deutschen Sprachbereich erschienenen Drucke* für das 16., 17. und 18. Jahrhundert publiziert – mittlerweile auch als Datenbanken. Dadurch entsteht eine retrospektive Nationalbibliographie. Für das 19. Jahrhundert gibt es solche Unternehmungen allerdings noch nicht; es bestehen derzeit nur entsprechende Sammelaufträge für zwei Bibliotheken: Für den Zeitraum 1801 bis 1870 ist dies die *Universitätsbibliothek Johann Christian Senckenberg* (Frankfurt am Main), für den Zeitraum 1870 bis 1912 ist dies die *Staatsbibliothek zu Berlin*. Beide weisen selbstverständlich die von ihnen im

³⁶ <http://www.weber-gesamtausgabe.de/correspDesc.xml> (letzter Aufruf 24.09.2018).

³⁷ Sarah Hartmann, *Entwicklungsprogramm für die GND 2017–2021. Gehalten auf der DINI AG Kompetenzzentrum Interoperable Metadaten WS 2017, Mannheim, 4. Mai 2017*, https://wiki.dnb.de/download/attachments/122262660/KIMWS2017_Tag1_Hartmann_GND.pdf?version=1&modificationDate=1494432064000&api=v2 (letzter Aufruf 14.9.2017).

Rahmen dieses Sammelauftrags erworbenen Publikationen im *OPAC* nach. Diese Nachweise sind aber bestandsorientiert; eine retrospektive Nationalbibliographie ist damit also nicht in Arbeit. Allerdings wäre ein solches, dezidiertes Verzeichnis – quasi ein *VD 19* – sehr wünschenswert.

Allen *Verzeichnissen der im deutschen Sprachbereich erschienenen Drucke* und dem Katalog der *DNB* (nicht der Normdatei!) ist gemeinsam, dass sie die tatsächlich gedruckten Ausgaben, nicht aber die Werke als Abstraktum verzeichnen. Allerdings ist im Bereich der (Brief-)Editionen selten bekannt, auf welche Ausgabe genau sich eine Erwähnung in der Quelle bezieht, wenn es denn mehrere Ausgaben gibt. Wichtiger ist daher das abstrakte Werk als die konkrete Fassung – oder gar einzelne Exemplare in den Bibliotheken – für eine projektübergreifende Identifizierung.³⁸ Normierte Werktitel sind zwar teilweise schon in der Gemeinsamen Normdatei vorhanden, aber eben längst nicht für jedes Werk. Wichtig wäre es, hier die GND zu erweitern und auch konsequent Verknüpfungen zwischen normiertem Werkdatensatz und Nachweisen in den *VDs* zu setzen.

Zusammenfassend ließe sich sagen, dass mit der Gemeinsamen Normdatei und den *VDs* schon viel Arbeit geleistet wurde und die notwendigen Infrastrukturen grundsätzlich schon bereitstehen. Im Hinblick auf den Einsatz in digitalen (Brief-)Editionen müssten sie allerdings optimiert, geöffnet und untereinander besser verknüpft werden.

Neben Personen, Orten und Werken werden in Briefen auch häufig Objekte oder Bauten erwähnt – insbesondere in Briefen von bzw. an Kunsthistoriker, Archäologen oder Naturwissenschaftler. Auch diese erwähnten Objekte gilt es projektübergreifend zu identifizieren. Dabei sind im Bereich der Archäologie und Kunstgeschichte schon entsprechende Normdateien oder vergleichbare, projektübergreifende Datenbanken entstanden, wie z. B. *iDAI.objects arachne*³⁹. Darüber hinaus sollten auch Museen ihre Bestände nach außen hin über Metadaten nachweisen und für die einzelnen Objekte URIs anbieten.

Eine Alternative zur Nutzung von Normdateien wurde mit dem Projekt *metagrid der Schweizerischen Akademie für Geistes- und Sozialwissenschaften* entwickelt.⁴⁰ Hier werden keine Normdatensätze angelegt, sondern lediglich Konkordanzen erstellt, die IDs zu Personen (und anderen Entitäten) aus den teilnehmenden Projekten aufeinander abbilden. D. h. *metagrid* hält selbst keine eigenen Metadaten vor, sondern vernetzt Metadaten aus den unterschiedlichen Projekten. Die Verknüpfung von verschiedenen Projekt-IDs geschieht dabei nicht ausschließlich manuell, sondern wird durch Prüfungswerkzeuge unterstützt.⁴¹ Vorteil ist, dass die Wissenschaftlerin nicht darauf angewiesen ist, dass es in einer

³⁸ Stadler (Anm. 30), 181–183.

³⁹ <https://arachne.dainst.org/> (letzter Aufruf 24.9.2017).

⁴⁰ <http://www.metagrid.ch> (letzter Aufruf 24.9.2017).

⁴¹ Sacha Zala/Christiane Sibille, „Metagrid und die Vernetzung von Metadaten“, in: *arbido* 3 (2017), <http://arbido.ch/de/ausgaben-artikel/2017/metadaten-datenqualitaet/C3%A4/metagrid> (letzter Aufruf 16.09.2017).

Normdatei einen entsprechenden Datensatz zu einer Person gibt, die man mit anderen Projekten vernetzen möchte. Zur Vernetzung muss dann aber zwingend die zentrale Instanz, d. h. *metagrid*, abgefragt werden. Daher stellt *metagrid* auch ein entsprechendes Javascript-Widget zur Verfügung, das die Daten in einer Web-Oberfläche, z. B. einer digitalen Edition, anzeigt. Bei der Benutzung von IDs aus Normdateien ist es dagegen möglich, dass sich digitale Publikationen untereinander vernetzen, ohne dass eine zentrale Instanz abgefragt werden muss.

4 Daten

Derzeit (Mitte September 2017) sind rund 24.500 edierte Briefe aus über 100 Publikationen im Webservice *correspSearch* nachgewiesen. Der Bestand wird laufend erweitert, indem Institutionen und Editionsprojekte – aber auch einzelne Wissenschaftlerinnen – Metadaten über edierte Briefe im CMI-Format online und unter einer freien Lizenz bereitstellen. In diesem Zusammenhang muss angemerkt werden, dass unter ‚edierten Briefen‘ nicht nur edierte Briefe im Wortsinne verstanden werden. Darunter würden dann ja nur editionswissenschaftlich erstellte Transkriptionen mit einer zumindest basalen Kommentierung fallen. Dabei fallen schon in Briefeditionen auch Daten über Briefe an, die gar nicht überliefert sind, sondern nur aus anderen Briefen oder weiteren Quellen (aus Tagebüchern, Briefjournalen etc.) belegt sind und ‚erschlossen‘ werden. Darüber hinaus gibt es auch Vorhaben, die Briefe mit Regesten erschließen (z. B. die Regesta-Ausgabe der Briefe an Goethe), oder Projekte, die lediglich Transkriptionen bereitstellen. Alle auf diese verschiedenen Arten wissenschaftlich aufbereiteten Briefe weist *correspSearch* nach.⁴² Demgegenüber weist *correspSearch* keine archivalischen Briefbestände nach – dies ist Aufgabe des Verbundkatalogs *Kalliope*. Wenngleich man aus den eben genannten Gründen eher von ‚wissenschaftlich aufbereiteten Briefen‘ sprechen müsste, wird der Einfachheit halber hier weiter von ‚edierten Briefen‘ gesprochen.

Edierte Briefe liegen in ganz verschiedenen Publikationstypen vor. Es ist naheliegend, dass viele Briefe aus digitalen Editionen nachgewiesen werden, die häufig zwar schwerpunktmäßig auch Briefe enthalten – aber nicht allein. Hierzu zählen beispielsweise die *Weber-Gesamtausgabe*, die Edition *Briefe und Texte aus dem intellektuellen Berlin* und die *edition humboldt digital*.⁴³ Reine digitale Briefeditionen, die Daten in *correspSearch* bereitstellen, sind z. B. die *Alfred-Escher-Briefedition* oder die *Briefe der Fruchtbringenden Gesellschaft und Beilagen*. Aber nicht nur Briefe aus digitalen, sondern auch aus gedruckten Editionen können bzw. werden bereits nachgewiesen. Darunter fallen zum einen Briefwechsel, wie

⁴²Wobei in den Daten und auf der Oberfläche noch keine Unterscheidung gemacht wird, das wird sich aber mit der Weiterentwicklung des *CMIF* und des Webservice ändern.

⁴³Berlin-Brandenburgische Akademie der Wissenschaften (Hg.), *edition humboldt digital*, Berlin, <http://edition-humboldt.de> (letzter Aufruf 24.9.2017).

etwa der Briefwechsel zwischen Karl August Varnhagen von Ense und Friedrich de la Motte-Fouqué, zum anderen gehören dazu Briefwechsel aus Gesamtausgaben, wie z. B. die *Kritische Gesamtausgabe* zu Friedrich Schleiermacher oder diejenige zu Samuel Thomas Soemmerring. Neben dezidierten Editionen enthalten aber auch andere Publikationen häufig edierte Briefe, wie z. B. Dissertationen oder Zeitschriftenaufsätze. Auch diese können in *correspSearch* nachgewiesen werden, wie es z. B. die *Open-Access-Zeitschrift HiN – Humboldt im Netz* handhabt.⁴⁴

Neben digitalen Editionen im engeren Sinne können auch gedruckte Editionen und andere Publikationen digital vorliegen – entweder als PDF oder als Digitalisat. Ersteres sollte in einem institutionellen Repositorium abgelegt werden und kann dann aus der bibliographischen Angabe heraus verlinkt werden – allerdings nur auf Ebene der Publikation, nicht auf Ebene der einzelnen Briefe. Anders sieht es mit älteren Editionen und Publikationen aus, die über die Digitalisierungssysteme der Bibliotheken bereitgestellt werden. Hier wird – spätestens bei der Ansicht im generischen *DFG-Viewer* – eine seitengenaue, permanente ID bereitgestellt, die als Verweisziel für den einzelnen Brief angegeben werden kann. Zwar könnte man darüber nachdenken, ob es möglich wäre, ein CMIF aus den von der Bibliothek erstellten Strukturdaten im METS/MODS-Format zu extrahieren. Dafür müssten aber bei Briefausgaben die Strukturdaten grundsätzlich so granular sein, dass sie jeden einzelnen Brief abbilden. Darüber hinaus enthält das CMIF mehr Daten, als in den Strukturdaten bisher notiert werden können (d. h. Norm-IDs, Orts- und Datumsangaben). Mit den geplanten Weiterentwicklungen des CMIF wird dieser Unterschied sehr wahrscheinlich noch größer werden.

Werden Daten aus verschiedenen Quellen aggregiert und unterliegen erst einmal keiner zentralen redaktionellen Kontrolle, kann es zu ‚Dubletten‘ kommen. Darunter fallen im Kontext von *correspSearch* Briefe, die mehrfach, aber in verschiedenen Editionen ediert vorliegen. Dieser Fall ist intendiert – im Idealfall kann man irgendwann alle Editionen eines Briefes über *correspSearch* abfragen und per Link aufrufen. Dennoch gibt es Nutzungsszenarien, wo man diese Mehrfachindizierungen filtern und dafür nur einen Datensatz erhalten möchte – etwa um die aggregierten Daten für eine (Korrespondenz-)Netzwerkanalyse zu nutzen. Um Briefe in *correspSearch* zu unterscheiden bzw. als identisch zu identifizieren, gibt es mehrere vorstellbare Ansätze. Der erste – und naheliegende – wäre, die archivalische Einheit anzugeben, auf die sich die Edition bezieht. Früher war das die Signatur, heute wäre es eine URI. Hier böten sich URIs aus dem Verbundkatalog *Kalliope* an, sofern sie denn den einzelnen Brief verzeichnen und nicht etwa ein Bündel von Briefen. Zweitens könnte man mit Hilfe von URIs explizit auf andere Editionen desselben Briefes verweisen, sofern diese URIs besitzen. Das ist insbesondere bei gedruckten Briefeditionen nicht der Fall. Hier könnte allerdings

⁴⁴Für eine komplette Übersicht und genaue bibliographische Angaben der in *correspSearch* aggregierten digitalen Briefverzeichnisse siehe <https://correspsearch.net/data.xq!?!=de> (letzter Aufruf 24.9.2017).

eine Bibliothek oder *correspSearch* selbst als URI-Anbieter stellvertretend einspringen. Drittens käme die Aufnahme von Incipits ins CMIF infrage, anhand welcher man einen Brief sehr wahrscheinlich eindeutig identifizieren könnte.

5 API

Wissenschaftlerinnen können den Datenbestand von *correspSearch* bequem über die Weboberfläche abfragen und sich anzeigen lassen. Die derzeit basalen Abfragemöglichkeiten werden nun im Rahmen des von der *DFG* geförderten Projektes weiterentwickelt.

Allerdings ist *correspSearch* nicht nur eine Website, sondern ein Webservice. Als solcher macht er die Daten nicht nur über eine Benutzeroberfläche zugänglich und durchsuchbar, sondern stellt sie auch zur automatisierten Abfrage über eine technische Schnittstelle, ein *Application Programming Interface (API)*, bereit. Dadurch wird *correspSearch* nicht zur Einbahnstraße, sondern im Gegenteil zum Umschlagplatz für die bereitgestellten Briefmetadaten.

Die aggregierten Daten können in verschiedenen Formaten abgerufen werden. Zuvorderst werden die Daten natürlich im selben Format angeboten, wie sie ursprünglich bereitgestellt wurden – nämlich als CMIF, d. h. TEI-XML. Sie können aber auch als CMIF in einer (derzeit noch in der Probephase befindlichen) JSON-Serialisierung abgerufen werden. Über diese beiden Schnittstellen hinaus bietet *correspSearch* die Daten zum einen noch als Texttabelle in CSV⁴⁵ an, zum anderen steht eine BEACON-Schnittstelle zur Verfügung, mit deren Hilfe Portale (wie z. B. die *Deutsche Biographie*) oder digitale Editionen auf die Nachweise edierter Briefe zu einer bestimmten Person automatisiert verlinken können.

Die Bereitstellung der aggregierten Briefmetadaten über eine API ermöglicht es, dass diese zur Analyse und Visualisierung von Korrespondenznetzwerken verwendet werden können. Der methodische Ansatz der Netzwerkanalyse, den Briefe (und deren Edition) geradezu evozieren, wurde – wie eingangs erwähnt – schon vielfach in der Forschung gewünscht. In Zukunft soll *correspSearch* zwar auch über eine eigene Visualisierungsoberfläche verfügen, um den Datenbestand leichter zugänglich zu machen. Häufig wird es aber notwendig sein, die aggregierten Daten (oder Teile davon) mit eigener Software und speziellen netzwerkanalytischen Methoden auszuwerten.

Darüber hinaus ermöglicht die API aber auch einen neuen Umgang mit den edierten Briefen. Wie eingangs schon angeführt, werden Briefe häufig im Rahmen von Briefwechseln, also der Korrespondenz zwischen i. d. R. zwei Personen, herausgegeben. Dadurch wird zwar der Briefwechsel gut bearbeitet und sichtbar, der Bezug zum übrigen Briefverkehr der zwei Protagonisten wird allerdings ausgeblendet bzw. nur durch die Konsultation anderer Editionen sichtbar. Ein gutes

⁴⁵Diese dienen weniger der technischen Weiterverarbeitung als vielmehr der einfachen Übernahme eines Suchergebnisses durch die Benutzerinnen der Weboberfläche.

Beispiel für einen solchen Fall ist Alexander von Humboldt: Bereits bei Arbeitsbeginn der *Alexander-von-Humboldt-Forschungsstelle* an der *Berliner Akademie der Wissenschaften* Anfang der 1960er Jahre hat man davon abgesehen, eine Gesamtausgabe zu unternehmen. Zu groß schien das Briefœuvre Humboldts zu sein, als dass dies möglich wäre. Man gab daher Teile der Humboldt'schen Korrespondenz in mehreren Briefausgaben heraus, die entweder auf bestimmte Themen oder eben Personen fokussierten. Mit Hilfe digitaler Briefverzeichnisse und dem Webservice *correspSearch* kann man nun die Humboldt'sche Korrespondenz wieder zusammenführen – aber nicht nur auf *correspSearch* selbst, sondern – dank der API – in Angeboten dritter, wie z. B. der Chronologie zu Alexander von Humboldts Leben in *edition humboldt digital*.

Mithilfe der *correspSearch*-API ist es aber auch möglich, die verteilt edierten Briefe direkt zu verknüpfen. Eine solche Vernetzungsfunktion ist ebenfalls in *edition humboldt digital* realisiert: Für jeden Brief wird *correspSearch* auf Briefe von und an Humboldt im selben Zeitraum hin abgefragt. Die Ergebnisse werden der Nutzerin direkt in *edition humboldt digital* unter dem Stichwort ‚Briefnetz erkunden‘ angezeigt. Perspektivisch⁴⁶ kann die Nutzerin sich dadurch wie in einer Gesamtausgabe bewegen und von Brief zu Brief springen, obwohl sie dabei im Regelfall die Edition wechseln wird. Darüber hinaus werden aber nicht nur die Briefe von und an Alexander von Humboldt aus anderen Editionen angezeigt, sondern auch diejenigen, die der jeweilige Korrespondenzpartner im selben Zeitraum empfangen oder versendet hat. Dieser ‚erweiterte Korrespondenzkontext‘ ist nicht unerheblich, kann eine Person doch über ein Ereignis, eine Publikation etc. an verschiedene Korrespondenzpartner schreiben – und das unter Umständen auch mit unterschiedlichem Inhalt. Denn wie die Forschung schon seit Längerem heraushebt, sind Briefe nicht nur subjektive Äußerungen von Individuen, sondern zudem noch stets stark auf den Adressaten bezogen. Die automatisierte Darstellung des ‚erweiterten Korrespondenzkontextes‘ enthebt die Wissenschaftlerin natürlich nicht von der inhaltlichen Auseinandersetzung mit den Briefen, unterstützt aber die Exploration der ‚benachbarten‘ Korrespondenz. Insofern hilft diese Funktion auch, ein methodisches Problem der gedruckten (und teilweise auch der digitalen) Briefedition zu überwinden, nämlich dass sich die Briefedition von der ‚Ordnungsfunktion Autor‘ lösen kann.⁴⁷

An diesem Beispiel wird deutlich, wie man verteilt vorliegende und mit unterschiedlichen Technologien realisierte Briefeditionen mithilfe entsprechender Schnittstellen tatsächlich vernetzen kann.

⁴⁶D. h. wenn alle edierten Briefe von und an A. v. Humboldt in *correspSearch* nachgewiesen sind.

⁴⁷Rüdiger Nutt-Kofoth, „Briefe herausgeben. Digitale Plattformen für Editionswissenschaftler und die Grundfragen der Briefedition“, in: Kristina Richts/Peter Stadler (Hg.), „*Ei, dem alten Herrn zoll' ich Achtung gern!*“. *Festschrift für Joachim Veit zum 60. Geburtstag*, München 2016, 575–586, hier: 584, <https://github.com/Edirom/Festschrift-Veit/releases> (letzter Aufruf 14.09.2017).

Im DFG-geförderten Projekt wird die API von *correspSearch* erweitert, u. a. auch um eine Schnittstelle, die die aggregierten Briefmetadaten als *Turtles* oder RDF-XML anbietet. Damit werden die Daten auch als *Linked Open Data* für Technologien des *Semantic Web* nutzbar. Diese Schnittstelle wird in Kooperation mit der *Digitalen Akademie der Akademie der Wissenschaften und der Literatur* (Mainz) entwickelt – auf Basis erster Erfahrungen aus einem Pilotprojekt.⁴⁸

6 *correspSearch* als Informationsinfrastruktur

Im Unterschied zu themen- oder forschungsspezifischen Briefdatenbanken ist der Webservice *correspSearch* als Informationsinfrastruktur konzipiert. Das bedeutet zum einen, dass er keine räumliche, zeitliche oder thematische Fokussierung hat und daher prinzipiell alle edierten Briefe nachweisen möchte. Zum anderen bedeutet dies, dass *correspSearch* einer fest umrissenen Aufgabe gewidmet ist und sich ansonsten in die Infrastrukturlandschaft einfügt. Das hat zur Folge, dass bestimmte Aspekte nicht von *correspSearch* abgedeckt werden sollen. So weist der Webservice nur edierte oder zumindest wissenschaftlich erschlossene Briefe nach – nicht aber deren archivalische Überlieferung. Letzteres ist Aufgabe des Verbundkatalogs *Kalliope*, zu dem mittlerweile sehr viele Archive des deutschsprachigen Raums beitragen. Ebenso wenig bietet *correspSearch* eine Personendatenbank an, um Absender und Empfänger identifizieren zu können. Für diese Aufgabe greift *correspSearch*, wie schon erwähnt, auf die GND oder andere Anbieter zurück. Die Vorhaltung gedruckter und digitaler Briefeditionen wiederum wird ebenfalls nicht durch *correspSearch* geleistet. Sie obliegt den jeweiligen Forschungseinrichtungen oder Bibliotheken. Die Liste ließe sich fortsetzen. Ziel ist jedenfalls, keine alles umfassende Datenbank zu bereitzustellen, sondern sich auf einen bestimmten Dienst zu fokussieren – eben den Nachweis edierter bzw. wissenschaftlich aufbereiteter Briefe.

Man könnte diesen Ansatz, den Joris van Zundert in ähnlicher Weise bereits 2012 vorgeschlagen hat, „Mikroinfrastruktur“ nennen, in Anlehnung an die in der Softwarearchitektur weit verbreiteten *Microservices*.⁴⁹ Diese übernehmen ganz spezielle Aufgaben und stellen die verarbeiteten Daten über definierte APIs zur Verfügung, wo sie wiederum von anderen *Microservices* weiterverwendet werden können. Ihre Vorteile sind – im übertragenen Sinne – auch die Vorteile

⁴⁸Max Grüntgens/Torsten Schrade, „Data repositories in the Humanities and the Semantic Web: modelling, linking, visualising“, in: Alessandro Adamou/Enrico Daga/Leif Isaksen (Hg.), *WHiSe 2016 Humanities in the Semantic Web. Proceedings of the 1st Workshop on Humanities in the Semantic Web (WHiSe)*, Aachen 2016, 53–64, <http://ceur-ws.org/Vol-1608/#paper-07> (letzter Aufruf 16.09.2017).

⁴⁹Joris van Zundert, „If You Build It, Will We Come? Large Scale Digital Infrastructures as a Dead End for Digital Humanities“, in: *Historical Social Research* 37/3 (2012), 165–186, hier: 176, 179–182.

einer Mikroinfrastruktur: Entwicklung und Wartung bleibt überschaubar, im Notfall kann eine Mikroinfrastruktur leichter ersetzt werden. Auch kann schneller ein erster Prototyp entwickelt werden, der bereits über basale Funktionen verfügt. So können Anforderungen und Anwendungsfälle unkomplizierter evaluiert werden. Gleichzeitig können sich leichter mehrere Akteure an der Errichtung einer Infrastrukturlandschaft beteiligen, da jeweils ein Dienst auf eine bestimmte Aufgabe fokussiert ist. Mikroinfrastrukturen können so unabhängig voneinander entwickelt und betrieben werden. Voraussetzung für eine ‚Landschaft der Mikroinfrastrukturen‘ ist allerdings, dass man sich über entsprechende Schnittstellen einigt, diese dann anbietet und andere wiederum nachnutzt. Dementsprechend soll der Webservice *correspSearch* im Rahmen der DFG-geförderten Weiterentwicklung auch um weitere Schnittstellen ergänzt werden, die den Dienst tiefer in die digitale Infrastrukturlandschaft integrieren.

Der Webservice *correspSearch* wird von einer Forschungseinrichtung entwickelt und dauerhaft vorgehalten. Diese Entwicklung ist im digitalen Zeitalter nicht neu: Informationsinfrastrukturen werden von verschiedenen Akteurstypen angeboten. Neben den Forschungseinrichtungen sind dies vor allem Bibliotheken, aber auch Universitäten und Archive. Größere Infrastrukturen werden meistens im Rahmen eines Kooperationsverbundes gestemmt. *correspSearch* wird als Mikroinfrastruktur zwar von einer einzelnen Forschungseinrichtung angeboten, dies geschieht aber in Kooperation mit Bibliotheken und Archiven – bzw. mit Rückgriff auf deren Dienste. Jedenfalls muss der Dialog zwischen den Akteuren weiter vertieft werden, um Dienste und deren Schnittstellen zu einer gemeinsamen, für die Wissenschaftscommunity sinnvollen Infrastrukturlandschaft verbinden zu können.

Literatur

- Appel, Bernhard/Breig, Werner/Buschmeier, Gabriele u.a., *Richtlinien-Empfehlungen zur Edition von Musikerbriefen*. Im Auftrag der Arbeitsgemeinschaft Musikerbriefe innerhalb der Fachgruppe Freie Forschungsinstitute in der Gesellschaft für Musikforschung, Mainz 1997.
- Baillot, Anne (Hg.), *Berliner Intellektuelle. Briefe und Texte aus dem intellektuellen Berlin um 1800*, Berlin, <http://www.berliner-intellektuelle.eu/> (letzter Aufruf 24.9.2017).
- Becker, Heinz, „Die Briefausgabe als Dokumenten-Biographie“, in: Wolfgang Frühwald/Hans-Joachim Mähl/Walter Müller-Seidel (Hg.), *Probleme der Brief-Edition. Kolloquium der Deutschen Forschungsgemeinschaft (DFG)*, Boppard 1977, 11–25.
- Berlin-Brandenburgische Akademie der Wissenschaften (Hg.), *edition humboldt digital*, Berlin, <http://edition-humboldt.de> (letzter Aufruf 24.9.2017).
- Berners-Lee, Tim, *Cool URIs Don't Change 1998*, <https://www.w3.org/Provider/Style/URI> (letzter Aufruf 24.9.2017).
- Berners-Lee, Tim/Fielding, Roy/Masinter, Larry, *RFC 3986. Uniform Resource Identifier (URI). Generic Syntax 2005*, <https://tools.ietf.org/html/rfc3986> (letzter Aufruf 24.9.2017).
- Bunzel, Wolfgang, „Briefnetzwerke der Romantik. Theorie – Praxis – Edition“, in: Anne Bohnenkamp/Elke Richter (Hg.), *Brief-Edition im digitalen Zeitalter*, Berlin/Boston 2013, 109–132.
- Bunzel, Wolfgang, „Briefe, Briefnetze, Briefnetzwerke. Überlegungen zur epistolaren Interkonnektivität“, in: *Fontanes Briefe ediert*, Würzburg 2014, 232–245.

- Dumont, Stefan, „Perspectives of the further development of the Correspondence Metadata Interchange Format (CMIF)“, in: *digiversity – Webmagazin für Informationstechnologie in den Geisteswissenschaften* (2015), <https://digiversity.net/2015/perspectives-of-the-further-development-of-the-correspondence-metadata-interchange-format-cmif/> (letzter Aufruf 16.9.2017).
- Eide, Øyvind, „Putting the Dialogue Back Together Re-Creating Structure in Letter Publishing“, in: *Computers and the Humanities* 37/1 (2003), 65–75.
- Elsaß, Stephan, *Sprachgeschichte von unten. Untersuchungen zum geschriebenen Alltagsdeutsch im 19. Jahrhundert*, Berlin/Boston 2011, DOI <https://doi.org/10.1515/9783110910568>.
- Grüntgens, Max/Schrade, Torsten, „Data repositories in the Humanities and the Semantic Web: modelling, linking, visualising“, in: Alessandro Adamou/Enrico Daga/Leif Isaksen (Hg.), *WHiSe 2016 Humanities in the Semantic Web. Proceedings of the 1st Workshop on Humanities in the Semantic Web (WHiSe)*, Aachen 2016, 53–64, <http://ceur-ws.org/Vol-1608/#paper-07> (letzter Aufruf 16.9.2017).
- Haaf, Susanne/Geyken, Alexander/Wiegand, Frank, „The DTA „Base Format“. A TEI Subset for the Compilation of a Large Reference Corpus of Printed Text from Multiple Sources“, in: *Journal of the Text Encoding Initiative* 8 (2015), DOI <https://doi.org/10.4000/jtei.1114>.
- Hartmann, Sarah, *Entwicklungsprogramm für die GND 2017–2021. Gehalten auf der DINI AG Kompetenzzentrum Interoperable Metadaten WS 2017, Mannheim, 4. Mai 2017*, https://wiki.dnb.de/download/attachments/122262660/KIMWS2017_Tag1_Hartmann_GND.pdf?version=1&modificationDate=1494432064000&api=v2 (letzter Aufruf 14.9.2017).
- Herres, Jürgen/Neuhaus, Manfred, „Vorwort der Herausgeber“, in: Jürgen Herres/Manfred Neuhaus (Hg.), *Politische Netzwerke durch Briefkommunikation. Briefkultur der politischen Oppositionsbewegungen und frühen Arbeiterbewegungen im 19. Jahrhundert*, Berlin 2002, 7–25.
- Mattenklott, Gert/Schlaffer, Hannelore/Schlaffer, Heinz, (Hg.), *Deutsche Briefe 1750–1950*, Frankfurt a.M. 1988.
- Motschmann, Uta, „Von den Registern“, in: Siegfried Scheibe/Waltraud Hagen/Christel Laufer u.a. (Hg.), *Vom Umgang mit Editionen. Eine Einführung in Verfahrensweisen und Methoden der Textologie*, Berlin 1988, 225–263.
- Nicksch, Reinhard M. G., *Brief*, Stuttgart 1991.
- Nutt-Kofoth, Rüdiger, „Briefe herausgeben. Digitale Plattformen für Editionswissenschaftler und die Grundfragen der Briefedition“, in: Kristina Richts/Peter Stadler (Hg.), „*Ei, dem alten Herrn zoll' ich Achtung gern*“. *Festschrift für Joachim Veit zum 60. Geburtstag*, München 2016, 575–586, <https://github.com/Edirom/Festschrift-Veit/releases> (letzter Aufruf 14.9.2017).
- Oellers, Norbert, „Wie sollten Briefwechsel ediert werden?“, in: Lothar Bluhm/Andreas Meier (Hg.), *Der Brief in Klassik und Romantik*, Würzburg 1993, 1–12.
- Rapp, Andrea, „Einige Anmerkungen zu Retrodigitalisierungs-Verfahren und Perspektiven digitaler Briefeditionen“, in: Peter Stadler/Joachim Veit (Hg.), *Digitale Edition zwischen Experiment und Standardisierung. Musik – Text – Codierung*, Tübingen 2009, 203–206.
- Schlawe, Fritz, *Die Briefsammlungen des 19. Jahrhunderts. Bibliographie d. Briefausg. u. Gesamtregister d. Briefschreiber u. Briefempfänger; 1815–1915*, Stuttgart 1969.
- Schmid, Irmtraud, „Anforderungen an die Kommentierung von Briefen und amtlichen Schriftstücken“, in: „*Ich an Dich*“. *Edition, Rezeption und Kommentierung von Briefen*, Innsbruck 2001, 35–45.
- Stadler, Peter, „Normdateien in der Edition“, in: *editio* 26 (2012), 174–183.
- Stadler, Peter, „Interoperabilität von digitalen Briefeditionen“, in: Hanna Delf von Wolzhagen (Hg.), *Fontanes Briefe ediert*, Würzburg 2014, 278–287.
- Stadler, Peter/Illetschko, Marcel/Seifert, Sabine, „Towards a Model for Encoding Correspondence in the TEI. Developing and Implementing <correspDesc>“, in: *Journal of the Text Encoding Initiative [Online]* 9 (2016), DOI <https://doi.org/10.4000/jtei.1433>.
- Strobel, Jochen, „Der Brief als Prozess. Entwurf und Konzept in der digitalen Edition“, in: Anne Bohnenkamp (Hg.), *Brief-Edition im digitalen Zeitalter*, Berlin/Boston 2013, 133–146.

- Strobel, Jochen, „Die Normierung von Metadaten als Standardisierungsinstrument in der digitalen Briefedition“, in: Michael Stolz/Yen-Chun Chen (Hg.), *Internationalität und Interdisziplinarität der Editionswissenschaft*, Berlin/Boston 2014, 99–107.
- TEI Correspondence SIG (Hg.), *Correspondence Metadata Interchange Format (CMIF) 2015*, <https://github.com/TEI-Correspondence-SIG/CMIF> (letzter Aufruf 16.9.2017).
- Voß, Jakob/Schindler, Matthias, *BEACON Link Dump Format 2014*, <http://gbv.github.io/beaconspec/beacon.html> (letzter Aufruf 16.6.2017).
- Weber, Jutta, „Briefnachlässe auf dem Wege zur elektronischen Publikation. Stationen neuer Beziehungen“, in: Anne Bohnenkamp/Elke Richter (Hg.), *Brief-Edition im digitalen Zeitalter*, Berlin/Boston 2013, 25–34.
- Woesler, Winfried, „Vorschläge für eine Normierung von Briefeditionen“, in: *editio* 2 (1988), 8–18.
- Zala, Sacha/Sibille, Christiane, „Metagrid und die Vernetzung von Metadaten“, in: *arbido* 3 (2017), <http://arbido.ch/de/ausgaben-artikel/2017/metadaten-datenqualit%C3%A4t/metagrid> (letzter Aufruf 16.9.2017).
- Zundert, Joris van, „If You Build It, Will We Come? Large Scale Digital Infrastructures as a Dead End for Digital Humanities“, in: *Historical Social Research* 37/3 (2012), 165–186.

Online-Ressourcen

Sämtliche Online-Ressourcen wurden letztmalig am 24.9.2017 eingesehen.

Arachne, <https://arachne.dainst.org/>.

BRIEFkasten, <http://www.textkritik.de/briefkasten/bkprojekt.htm>.

Carl-Maria-von-Weber-Gesamtausgabe, <http://www.weber-gesamtausgabe.de/>.

correspSearch, <https://correspsearch.net/data.xql?l=de>.

Cultures of Knowledge. The fourth phase, <http://www.culturesofknowledge.org/?p=8455>.

Early Modern Letters Online, <http://emlo.bodleian.ox.ac.uk/>.

Frühneuzeitliche Ärztebriefe des deutschsprachigen Raums (1500–1700), <http://www.medizingeschichte.uni-wuerzburg.de/akademie/index.html>.

Metagrid, <http://www.metagrid.ch>.

Reassembling the Republic of Letters, <http://www.republicofletters.net/>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Netzliteratur als Soziales Medium. Zur Kritik des Werkbegriffs

Thomas Ernst

1 Vom literarischen Werk zum literarischen Netzwerk? Zur Einführung in ein Problem der Netzliteraturwissenschaft

Der digitale Medienwandel und insbesondere die Etablierung Sozialer Medien bringen neue literarische Phänomene und einen Wandel der literarischen Kommunikation mit sich. Wenn eine *Digitale Literaturwissenschaft* auch diese Entwicklungen zu ihrem Gegenstandsbereich zählt, muss sie sich fragen, inwiefern sie bestehende Konzepte, Begriffe und Methoden der Literaturwissenschaft adaptieren kann oder modifizieren muss, um diesem Wandel gerecht zu werden – zumal die Buchwissenschaft noch 2015 als Desiderat markiert, dass die „Erforschung [...] der Vernetzung von Lesemedien und Lesern [...] bisher in keiner angemessenen Relation zur Bedeutung von Digitalität und Vernetzung im

Dieser Beitrag deckt sich zu großen Teilen mit dem ursprünglich 2017 zur Konferenz eingereichten Beitrag. Auf Basis der Präsentationen und Diskussionen in der *Villa Vigoni* habe ich allerdings 2018 Ergänzungen und begriffliche Modifikationen vorgenommen. Insbesondere geht es dabei um eine stärkere Nutzung des ‚Netzwerk‘-Begriffs (so wurde u. a. im gekürzten Beitragstitel ‚Digitale Literatur‘ durch ‚Netzliteratur‘ ersetzt), um die Ersetzung des ‚offenen Werkbegriffs‘ durch einen ‚dynamischen Werkbegriff‘ und um eine stärkere Relativierung der Werk- und Textbegriffe von Roland Barthes in meinen Ausführungen. Zudem habe ich den Vorschlag umgesetzt, das dritte Kapitel um Ausführungen zur quantitativen Begriffsnutzung zu erweitern. Ich danke Julia Nantke für ihren Kommentar zu meinem Beitrag sowie den Diskutant*innen für ihre Hinweise. Zudem danke ich Annette Gilbert und Thomas Kater für unseren Austausch zum Werkbegriff.

T. Ernst (✉)
University of Antwerp, Antwerp, Belgium
E-Mail: Thomas.Ernst@uantwerpen.be

Alltag der Menschen steht“.¹ Auf dem DFG-Symposium 2017 *Digitale Literaturwissenschaft* haben sich – neben dem vorliegenden Beitrag – noch drei weitere Beiträge mit literarischen bzw. medialen Produktionen beschäftigt, die ohne eine vernetzte digitale Kommunikation nicht funktionieren würden. Dabei wurden u. a. das kollektive Schreibprojekt *Tausend Tode Schreiben* (Julia Nantke), eine offene Annotationsplattform populärer Lieder namens *genius.com* (Alexander Nebrig) sowie Modeblogs wie *journelles.de* oder *cestclairette.com* (Jörg Schuster) als mögliche Gegenstände einer *Digitalen Literaturwissenschaft* reflektiert.²

Die Präsentation und Diskussion dieser vier Beiträge legen nahe, dass sich unter den Verhältnissen einer literarischen Online-Kommunikation Fragestellungen ergeben, mit denen sich eine noch zu konstituierende ‚Netzliteraturwissenschaft‘ als Teilbereich der *Digitalen Literaturwissenschaft* befassen müsste.³ Welche Begriffe und Konzepte helfen bei der Analyse der interaktiven bzw. kollaborativen netzliterarischen Produktionen? Inwiefern verändern sich die Beziehungen von Autor*innen und Leser*innen? An welche literaturhistorischen Traditionslinien schließen diese Phänomene an und mit welchen etablierten literaturwissenschaftlichen Begriffen und Modellen lassen sie sich angemessen beschreiben?

Diesen Fragen will sich dieser Beitrag zuwenden, indem er nach dem Verhältnis literaturwissenschaftlicher Werkbegriffe zur vernetzten literarischen Online-Kommunikation fragt. Ein *Symposium zur Gegenwart des literarischen Werkbegriffs* stand 2015 im Schloss Herrenhausen unter dem Titel *Wiederkehr des Werks?: Während einzelne Vorträge für die Restauration des Werkbegriffs als eines zentralen Begriffs der Literaturwissenschaft plädierten* (u. a. Werner Wolf),

¹Axel Kuhn/Susanne Kraus, „Nutzergenerierte Texte in digitalen Netzwerken“, in: Ursula Rautenberg/Ute Schneider (Hg.), *Lesen. Ein interdisziplinäres Handbuch*, Berlin/Boston 2015, 679–699, hier: 696.

²Vgl. dazu die Beiträge von Julia Nantke, Alexander Nebrig und Jörg Schuster aus diesem Band. Einen Eindruck von der Unterschiedlichkeit der genannten Gegenstände erhält man unter: <https://orbanism.com/1000tode/>; <https://genius.com/>; <https://www.journelles.de/>; <http://cestclairette.com/> (letzter Aufruf der Links 1.5.2018).

³Eine ‚Netzliteraturwissenschaft‘ müsste direkt an die Arbeiten von u. a. Christiane Heibach, Peter Gendolla und Jörgen Schäfer, Roberto Simanowski oder Simone Winko zu diesem Themenfeld anschließen; auch viele Forschungsprojekte im Göttinger DFG-Graduiertenkolleg *Literatur und Literaturvermittlung im Zeitalter der Digitalisierung* (<https://www.uni-goettingen.de/de/kollegiat/innen+und+forschungsstudent/innen/447610.html> [letzter Aufruf 1.5.2018]) oder Veröffentlichungen wie u. a. Maren Conrad/Theresa Schmidtke/Martin Stobbe (Hg.), *Textpraxis. Digitales Journal für Philologie* 8/2 (2017), Sonderausgabe #2: *Digitale Kontexte. Literatur und Computerspiel in der Gesellschaft der Gegenwart*; oder Innokentij Kreknin/Chantal Marquardt (Hg.), *Textpraxis. Digitales Journal für Philologie* 7/2 (2016), Sonderausgabe #1: *Das digitalisierte Subjekt. Grenzbereiche zwischen Fiktion und Alltagswirklichkeit* tragen zu dieser Teildisziplin bei; vgl. auch die gesammelten Beiträge in Johannes Auer/Christiane Heibach/Beat Suter (Hg.), *netzliteratur.net. Netzliteratur // Internetliteratur // Netzkunst* (2018), <https://www.netzliteratur.net> (letzter Aufruf 1.5.2018), die sich mit den Grundfragen einer solchen ‚Netzliteraturwissenschaft‘ befassen. Ich habe in meiner Habilitationsschrift eine ‚Netzliteraturwissenschaft‘ zu begründen versucht, vgl. Thomas Ernst, *Literatur als soziales Medium in der digitalisierten Gesellschaft. Zur Begründung einer Netzliteraturwissenschaft*, Essen 2018 (unveröff. Habilitationsschrift; wird 2022/2023 veröffentlicht).

interessierten sich andere für die verschiedenen Praxen und Akteure, die bei der Zuschreibung der Werkförmigkeit an literarische Gegenstände eine wichtige Rolle spielen.⁴ Für das Verhältnis des Werkbegriffs zur Online-Welt zeigte Alexander Starre, „dass das digitale Zeitalter den Status des ‚Werks‘ und das Beschreibungsparadigma ‚Werk‘ destruiere, indem auch mit flüchtigem Text und daher mit einer Ästhetik des Augenblicks gearbeitet wird. Dies führe zu einer Neuverhandlung der am Werk beteiligten Akteure.“⁵

Dieser allgemeinen Beobachtung will der vorliegende Beitrag genauer nachspüren, indem er von der Beobachtung ausgeht, dass sich durch den Wandel vom gedruckten Buch zum digitalen Text und durch die Potenziale der vernetzten Online-Kommunikation nicht nur die mediale Form der Literatur, sondern auch die literarische Kommunikation zwischen Autor(texten) und Leser(texten) nachhaltig verändert. Es kann gezeigt werden, dass sich die noch immer wirksame Vorstellung eines statischen Werks, das durch die Autorintention definiert und schriftlich fixiert worden ist, für den Bereich der Netzliteratur als problematisch erweist. Im Gegensatz dazu hilft ein dynamischer Werkbegriff, die neuen Formen der netzliterarischen Kommunikation und insbesondere die Relevanz der Benutzerschnittstellen adäquat zu beschreiben.

Dabei nimmt dieser Beitrag eine diskursanalytische Perspektive ein – Foucaults Beschreibung der Autorfunktion untersucht in ähnlicher Weise auch die diskursive Konstruktion des Werkbegriffs.⁶ Dieser Ansatz interessiert sich nicht für einen ontologischen oder materiellen Status des Werkbegriffs, sondern für seine strukturelle Funktion in einem diskursiven Ordnungsgefüge: ‚Werke‘ erscheinen diskursanalytisch als „machtimprägnierte, künstlich-kunstvoll hergestellte disperse Einheiten, die sich wesentlich aus Differenzen ergeben“. Die Kategorie ‚Werk‘ verfolgt eine „aktive Sinnordnungspolitik“ und wirkt disziplinierend:

⁴Vgl. u. a. die Vorträge von Andrea Albrecht und Steffen Martus auf der Konferenz, entsprechend der Zusammenfassung von Thomas Kater; vgl. Thomas Kater, „Keine Wiederkehr des Dagewesenen. Zu Gegenwart und Potential des literarischen Werkbegriffs“ (Conference Proceedings of: „Wiederkehr des Werks? Symposium zur Gegenwart des literarischen Werkbegriffs“. Schloss Herrenhausen, Hannover, 21.–23. Oktober 2015.), in: *JLOnline – Journal of Literary Theory* (8.3.2016), <http://www.jltonline.de/index.php/conferences/article/view/783/1838> (letzter Aufruf 1.5.2018), URN nbn:de:0222-003213; vgl. Elisabeth Weiß, „Wiederkehr des Werks? Zur Gegenwart des literarischen Werkbegriffs“ (Symposium in Hannover v. 21.–23. Oktober 2015), in: *Zeitschrift für Germanistik. Neue Folge* 26/2 (2016), 420–423. Der Band zur Konferenz soll 2019 erscheinen: Lutz Danneberg/Annette Gilbert/Carlos Spørhase (Hg.), *Das Werk. Zum Verschwinden und Fortwirken eines Grundbegriffs*, Berlin 2019 (im Erscheinen).

⁵Starre nach: Weiß (Anm. 4), 421.

⁶Vgl. Michel Foucault, „Was ist ein Autor?“, in: Ders.: *Schriften zur Literatur*, aus dem Französischen von Karin von Hofer und Anneliese Botond, Frankfurt a. M. 1988 [frz. 1969], 7–31; Michel Foucault, *Die Ordnung des Diskurses*, mit einem Essay von Ralf Konersmann, Frankfurt a. M. 1993 [frz. 1972]; Peter V. Zima, „Anwesenheit und Abwesenheit des Werks. Zu Foucaults Subjekt- und Werkbegriff“, in: Klaus-Michael Bogdal/Achim Geisenhanslüke (Hg.), *Die Abwesenheit des Werkes. Nach Foucault*, Heidelberg 2006, 181–191.

„Die Diener des Werks sind zugleich die Herren der Interpretation.“⁷ Aus einer kritischen Perspektive rücken daher – neben dem literarästhetischen Diskurs – insbesondere die Diskurse des digitalen Publizierens und der Literatur in Sozialen Medien, der literatur- und buchwissenschaftliche sowie der juristische Diskurs in den Mittelpunkt, denn Begriffe wie ‚Werk‘ oder ‚geistiges Eigentum‘ nehmen in verschiedenen Diskursen unterschiedliche Bedeutungen an, die dann wiederum in literarischen oder philosophischen Texten interdiskursiv vermittelt werden.⁸

Indem dieser Beitrag auch den juristischen Diskurs reflektiert, verweist er auf die gesellschaftspolitische Dimension des digitalen Wandels und auf die Notwendigkeit, Begriffe und Modelle, die als Dispositive spezifische Literaturkonzepte, geisteswissenschaftliche Schulen und literaturbetriebliche Geschäftsmodelle stützen, zu problematisieren. Dies scheint umso wichtiger, als breit wahrgenommene literatur- und medienwissenschaftliche Arbeiten zum Themenfeld dieses Beitrags entweder für die Bewahrung traditioneller Konzepte plädieren, wie zum Beispiel jenes des ‚geistigen Eigentums‘, und dieses in scharfer Abgrenzung von den digitalen Produktions- und Rezeptionsformen verteidigen,⁹ oder konstatieren, dass gegenwärtig noch keine angemessenen Aussagen über die Digitalisierung getroffen werden könnten.¹⁰ Dass eine Netzliteraturwissenschaft sich zu diesen Debatten anders positionieren muss, scheint evident. Dieser Beitrag will eine solche Position für den Werkbegriff entwickeln.

2 Werk, Text, Netzwerk: Statische und dynamische Konzeptionen des literarischen Werkbegriffs

Der Werkbegriff kann nur sinnvoll problematisiert werden, wenn er in seiner gewachsenen Bedeutung differenziert wird. Dazu werde ich zunächst in einem kursorischen historischen Rückgriff seine antike Begriffsbildung und Konzeptionen eines geschlossenen, statischen Werks beschreiben sowie in einem zweiten Schritt literaturwissenschaftliche Entwürfe eines dynamischen und pragmatischen Werkbegriffs.

⁷ Harro Müller, „Einige Notizen zu Diskurstheorie und Werkbegriff“, in: Jürgen Fohrmann/Harro Müller (Hg.), *Diskurstheorien und Literaturwissenschaft*, Frankfurt a. M. 1988, 235–243, hier: 239.

⁸ Vgl. Jürgen Link/Ursula Link-Heer, „Diskurs/Interdiskurs und Literaturanalyse“, in: *Zeitschrift für Literaturwissenschaft und Linguistik* 20 (1990), 88–99; Rolf Parr, „Interdiskurstheorie/Interdiskursanalyse“, in: Ders./Clemens Kammler/Ulrich J. Schneider (Hg.), *Foucault-Handbuch. Leben, Werk, Wirkung*, Stuttgart 2008, 202–206; Jürgen Link/Rolf Parr, „Semiotik und Interdiskursanalyse“, in: Klaus-Michael Bogdal (Hg.), *Neue Literaturtheorien. Eine Einführung*, Göttingen 2005, 108–133.

⁹ Philipp Theisohn, *Literarisches Eigentum. Zur Ethik geistiger Arbeit im digitalen Zeitalter. Essay*, Stuttgart 2012, 121–124.

¹⁰ Monika Dommann, *Autoren und Apparate. Die Geschichte des Copyrights im Medienwandel*, Frankfurt a. M. 2014, 296.

2.1 Fixierte Druckwerke, das Werk als Produkt und die Werkherrschaft der Autor*innen: Statische Werkbegriffe

Etymologisch leitet sich der Werkbegriff aus dem altgriechischen *ἔργον* (*érgon*) ab, er bezeichnet das Ergebnis einer produktiven Tätigkeit. In der Rhetorik setzt die Geschichte des heutigen Werkbegriffs mit der geformten mündlichen Rede ein, die in Ereignis und Rezeption unterteilt wird: „Die Werkhaftigkeit der Rede manifestiert sich primär nicht in einem die Zeit überdauernden Objekt, sondern kommt im Vortrag [...] bzw. in der durch ihn bewirkten Anschlußhandlung des Publikums zum Ausdruck.“¹¹ In Platons *Phaidros* findet sich das berühmte Zwiegespräch über die Redekunst zwischen Sokrates und Phaidros, in dem Sokrates begründet, weshalb die Redekunst unbedingt den Schriftwerken vorzuziehen sei: Die Erscheinungsweise und Qualität eines Werks wird von Platon in Relation zu ihrer medialen Form und zu seiner Wirkung auf das Publikum bewertet.¹²

Für Martin Kölbel wiederum setzt die Geschichte des Werkbegriffs mit der *Eudemischen Ethik* von Aristoteles an. Er verweist auf Aristoteles' fundamentale Differenzierung aller menschlichen Handlungen nach Zweck oder Ziel, analog dazu werde das Werk von Aristoteles als eine Tätigkeit oder aber als Ergebnis einer Handlung bestimmt.¹³ Hier ist bereits die später in der Reflexion der literarischen Kommunikation wichtige Unterscheidung von Produktionsprozess und dem nach dem Plazet des Autors bzw. der Autorin veröffentlichten Text angelegt, allerdings habe sich als „wirkungsmächtige Denotation“ des Werkbegriffs „die eines selbstständigen, dingartigen Produkts [durchgesetzt], das durch seine materielle Beschaffenheit von seinen Produktionsbedingungen freigestellt ist.“¹⁴

In dieser Traditionslinie steht die Vorstellung eines „Erzeugnis[ses] mit einer einmaligen, unantastbaren und endgültig fixierten Gestalt“, die sich allerdings erst deutlich später etabliert, Ostermann setzt dies im italienischen Humanismus des 14. Jahrhunderts an.¹⁵ Ein solches Verständnis des literarischen Werkbegriffs konzentriert sich auf das Produkt schriftstellerischer Tätigkeit und versteht dieses weniger als Teil einer literarischen Kommunikation denn vielmehr als eigenständige, entrückte und also statische Entität. So konstatiert die *Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*, beim literarischen Werk handele es sich um

¹¹ Eberhard Ostermann, „Werk“, in: Gerd Ueding (Hg.), *Historisches Wörterbuch der Rhetorik. Band 9: St-Z*, Tübingen 2009, 1352–1362, hier: 1352.

¹² Vgl. Platon, „Phaidros“, in: Ders., *Sämtliche Werke*, hg. von Erich Loewenthal, Bd. 2, Berlin/Heidelberg 1940 (gr. um 370 v. u. Z.), 411–482, <http://www.zeno.org/nid/20009262660> (letzter Aufruf 1.5.2018).

¹³ Vgl. Martin Kölbel, „Das literarische Werk. Zur Geschichte eines Grundbegriffs der Literaturtheorie“, in: *Text.Kritische Beiträge* 10 (2005), 27–44, hier: 28–32.

¹⁴ Ebd., 32.

¹⁵ Ostermann (Anm. 11), 1355.

das fertige und abgeschlossene Ergebnis der literarischen Produktion, das einem Autor zugehört und in fixierter, die Zeit überdauernder Form vorliegt, so daß es dem Zugriff des Produzenten ebenso entzogen ist wie dem Verbrauch durch den Rezipienten.¹⁶

Der literarische Werkbegriff entwickelt in dieser Lesart eine Nähe zum Werkbegriff im Militärwesen, dort meint das ‚Werk‘ den mit Wällen und Gräben befestigten geschlossenen äußeren Teil einer Festung.¹⁷

Die Einführung des Buchdrucks im 15. Jahrhundert verändert die mediale Gestalt, die Distributions- und die Archivierungsformen literarischer Werke nachhaltig: Ihre Schriftkörper werden grundlegend anders gestaltet, die Differenzierung zwischen (Autor-)Handschrift und (Verlags-)Druckversion definiert sowohl die Autorintention zur Veröffentlichung als auch die Warenförmigkeit des Druckwerks. Bis heute stellt sich für gedruckte literarische Bücher daher die Frage, inwiefern die Autor*innen auch die materiale Gestaltung des Buches – u. a. seine Ausstattung oder Typographie – bestimmt haben: Alexander Starre schlägt die Bezeichnungen ‚Buchwerk‘ und ‚bibliographischer Autor‘ für solche Fälle vor,¹⁸ Annika Rockenberger und Per Röcken rücken aus editionsphilologischer Sicht den „material text“ in den Vordergrund, dessen Materialität umfasse „alle chemischen und/oder im weitesten Sinne physikalischen Eigenschaften des Dokuments und der Schrift, einschließlich graphisch-visueller Eigenschaften der Schriftzeichen“.¹⁹ Die Autorschaft eines Werks konstituiert sich aus dieser Sicht nicht nur durch die Produktion einer sprachlichen Form und der Inhalte, sondern auch durch die materielle Gestaltung eines literarischen Werks.

Die Autorschaft eines Werks gewinnt Ende des 18. Jahrhunderts eine neue Bedeutung, denn das literarische Werk wird in einer Ordnungsstruktur positioniert, in der Autorname, Titel sowie Erscheinungsort und -jahr zentral stehen.²⁰ Die das Werk prägende und somit gestärkte Rolle des Autors, die in der Rechtswissenschaft als „urheberrechtliche Prägetheorie“ firmiert,²¹ wird zu einer zentralen Figur bei der Etablierung des modernen Literaturbetriebs mit seinen Instanzen und Institutionen: Das sind vor allem die Verlage und der Buchhandel. Deren Geschäftsmodelle basieren auf dieser spezifischen Ökonomie des Autor-Werk-Verhältnisses, die ab

¹⁶Horst Thomé, „Werk“, in: *Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*. Bd. III: P-Z, hg. von Jan-Dirk Müller, gem. mit Georg Braungart u. a., Berlin/New York 2003, 832–834, hier: 832.

¹⁷Vgl. Brockhaus, *Brockhaus Enzyklopädie*. Band 29: VERTI-WETY, Mannheim 2006, 748.

¹⁸Vgl. Kater (Anm. 4), 5.

¹⁹Annika Rockenberger/Per Röcken, „Wie ‚bedeutet‘ ein ‚material text‘?“, in: Wolfgang Lukas/Rüdiger Nutt-Kofuth/Madleen Podewski (Hg.), *Text – Material – Medium. Zur Relevanz editorischer Dokumentationen für die literaturwissenschaftliche Interpretation* (Beihefte zu *editio* 37), Berlin/Boston 2014, 25–52, hier: 41.

²⁰Vgl. Ostermann (Anm.11), 1356; Foucault 1988 (Anm. 6), 18 f.

²¹Vgl. Malek Barudi, *Autor und Werk – eine prägende Beziehung. Die urheberrechtliche Prägetheorie im Spiegel der Literaturwissenschaft*, Tübingen 2013, 19–28.

den 1770er Jahren in literarischen und philosophischen Diskursen intensiv diskutiert und ab den 1830er Jahren im juristischen Diskurs als ‚Urheberrecht‘ gefasst wird: Hier wird zwischen der Gedankenform, der *geistigen* Ausgangsidee eines Werkes, die eine Autorin in sich trägt, und ihrer *Materialisierung im gedruckten Werk* unterschieden. Diese besondere Qualität, ein *eigentümliches Werk* – also einen besonderen ästhetischen Gegenstand – *in sich* gedacht zu haben, wird zum Begründungsmoment des ‚geistigen Eigentums‘, das der Autor unabdingbar an seinem Werk behält und das ein „Zeitalter der Werkherrschaft“²² durch die Autor*innen begründet. Streng genommen ist dieses Werk als Idee bereits im Geiste des Autors bzw. der Autorin innerlich präsent, ohne überhaupt veräußert zu sein, weshalb „die gegenständliche Ausführung des Werkes, das sichtbare Original, bereits die erste Vervielfältigung“²³ ist. Zwar kann die Autorin die Nutzungs- und Verlagsrechte an ihrem Werk veräußern, die geistigen Eigentumsrechte an ihrer ‚Schöpfung‘ verbleiben jedoch bei ihr, weshalb sie auch den Nachdruck regulieren kann.

Diese Autor-Werk-Ökonomie wirkt bis heute – als Urheber-Schöpfung-Verhältnis – im deutschen Urheberrecht fort, das sich auf den Schutz der Urheber konzentriert (§ 1), die als „Schöpfer des Werkes“ (§ 7) bestimmt werden. Der Werkbegriff ist dabei sowohl auf den persönlichen Schöpfungsakt als auch auf eine Mindestqualität (‚Schöpfungshöhe‘) des Gegenstands ausgerichtet: „Werke im Sinne dieses Gesetzes sind nur persönliche geistige Schöpfungen.“²⁴ Zu beachten ist dabei, dass die Schöpfungshöhe im Regelfall bereits von Werken „mit einer geringen Gestaltungsintensität“ und von besonders kurzen Texten wie Werbeslogans erfüllt sein kann.²⁵ Ein solcher Begriff des literarischen Werkes als fixiertes, enthobenes und geschütztes Produkt eines (individuellen) Schöpfungsprozesses, das der intentionalen Werkherrschaft der Autoren unterworfen ist und sich in der medialen Form des fixierten Druckwerkes zeigt, legitimiert somit zugleich die heute vorherrschenden juristischen und buchökonomischen Rechts- und Geschäftsmodelle.

Auch einige wirkungsmächtige editionsphilologische Ansätze korrelieren mit diesem auf die Autorintention und die drucktechnische Fixierung ausgerichteten Werkbegriff. Gunther Martens bezieht sich in seiner editionsphilologischen Beantwortung der Frage, wie das literarische Werk – als Gesammeltes Werk, Gesamtwerk oder einzelnes Werk – zu bestimmen sei, auf die Autorintention. Über die Umwandlung eines Texts in ein Werk bestimme die Autorin, denn die

²² Heinrich Bosse, *Autorschaft ist Werkherrschaft. Über die Entstehung des Urheberrechts aus dem Geist der Goethezeit*, Paderborn u. a. 1981, 103.

²³ Ekkehard Gerstenberg, *Die Urheberrechte an Werken der Architektur und der Photographie*, München 1968, 60.

²⁴ Bundesministerium der Justiz und für Verbraucherschutz, *Gesetz über Urheberrecht und verwandte Schutzrechte* (2018), § 2.2, <https://www.gesetze-im-internet.de/urhrg/> (letzter Aufruf 1.5.2018).

²⁵ Vgl. Richard Hahn/Martin Schippan, *Rechtsfragen im Verlag. Urheberrecht, Verlagsrecht & Co*, Berlin/Boston 2014, 6 u. 9.

Textgenese sei nur eine Vorstufe des Werks. Die Einheit des Werks jedoch „legt allein durch seinen Entschluß, einen Text zu veröffentlichen, der Autor fest.“ Ein solchermaßen konstituiertes Werk sei vor allem bestimmt durch seine „Grenze“, denn es „setzt in seinem Begrenzt-Sein uns als Interpreten und Editoren Grenzen.“²⁶ Carlos Spoerhase bezeichnet den „Eintritt des Textes in die soziale Ordnung“²⁷ allgemein als *Werksozialität*. Die Werkherrschaft der Autor in beginnt bei ihm dort, „wo der Autor dem Interpreten hermeneutisch relevantes Material vorenthalten kann“²⁸ und endet dort, „wo der Kommunikationsakt geschlossen wird“ und die Autorin der Leserin kein zusätzliches Material mehr bereitstellen kann, letztlich also durch die nach dem Ableben des Autors ausgelöste „endgültige Versiegelung des Werks“.²⁹

2.2 Schreiben und Lesen als Prozess. Literaturwissenschaftliche Entwürfe eines dynamischen Werkbegriffs

Daneben etablieren sich auch eher pragmatische literaturwissenschaftliche Entwürfe eines dynamischen Werkbegriffs. In der Moderne werden die Abgeschlossenheit des einzelnen Werks und seine Enthobenheit durch Theorien der Intertextualität oder der durch Leser zu füllenden Leerstellen relativiert und sein prinzipiell offener Charakter, seine Unbestimmtheit, fokussiert.³⁰ Bei der Auseinandersetzung mit diesen in der Philosophie, der Literaturwissenschaft und der Editionsphilologie entwickelten Versuchen, einen pragmatischen oder dynamischen Werkbegriff zu entwerfen, helfen (editions-)philologische Differenzierungen des Werkbegriffs sowie die Nutzung von Para-, Kon- und Intertexten zur Bestimmung eines Werks sowie von populärkulturellen und post-strukturalistischen Ansätzen zur Relativierung des statischen Werkbegriffs.

Zunächst muss der Werkbegriff binnendifferenziert werden, da er verschiedene Ebenen besitzt. Spoerhase hat 2007 vorgeschlagen, fünf Ebenen des Werkbegriffs

²⁶ Gunther Martens, „Das Werk als Grenze. Ein Versuch zur terminologischen Bestimmung eines editorischen Begriffs“, in: *editio* 18 (2004), 175–186, hier: 185.

²⁷ Carlos Spoerhase, „Was ist ein Werk? Über philologische Werkfunktionen“, in: *Scientia Poetica. Jahrbuch für Geschichte der Literatur und der Wissenschaften/Yearbook for the History of Literature, Humanities, and Sciences* 11 (2007), 276–344, hier: 329.

²⁸ Ebd., 339.

²⁹ Ebd., 341.

³⁰ Vgl. u. a. Gérard Genette, *Palimpseste. Die Literatur auf zweiter Stufe*, Frankfurt a. M. 1993 [frz. 1982]; Wolfgang Iser, *Die Appellstruktur der Texte. Unbestimmtheit als Wirkungsbedingung literarischer Prosa*, Konstanz 1970; Julia Kristeva, „Bachtin, das Wort, der Dialog und der Roman“, in: Jens Ihwe (Hg.), *Literaturwissenschaft und Linguistik. Ergebnisse und Perspektiven*, Bd. 3: *Zur linguistischen Basis der Literaturwissenschaft II*, Frankfurt a. M. 1972, 345–375 [frz. 1967].

zu unterscheiden: *erstens* das Gesamtwerk, also alle (noch ungeordneten) textlichen Hinterlassenschaften eines Autors oder einer Autorin, die als *Überlieferung* (*Patrimonium*) zu bezeichnen wären; *zweitens* der als veröffentlichbares Ausgangsmaterial hieraus produzierte *Text*, der „in einer verbindlichen Leseordnung“ sortiert worden ist. Dieser Text kann dann *drittens* durch vier Kriterien zum *Werk* (*Opus*) qualifiziert werden:³¹ durch „einen Titel (paratextuelles Kriterium), [...] einen Veröffentlichungsakt (institutionelles Kriterium), [...] die Autorabsicht (intentionales Kriterium) oder [...] seinen Geschlossenheitsgrad (Anfang und Ende) bzw. Vollendungsgrad (ästhetisches Kriterium).“³² Schließlich gibt es *viertens* von einem *Opus* eine bestimmte *Ausgabe* (*Euvre*) sowie *fünftens* von einem spezifischen *Euvre* ein jeweiliges *Exemplar*.³³ Diese Differenzierung ist zwar vor allem aus editionsphilologischen Grundannahmen abgeleitet worden und privilegiert problematische Größen wie die Autorabsicht oder den Geschlossenheitsgrad, kann jedoch gerade deshalb als Folie für die spätere Problematisierung eines Werkbegriffs genutzt werden, der sich aus der Bestimmung des Werks als *Opus* ableitet und literarästhetisch wie buchökonomisch und juristisch folgenreich ist.

Eine breitere Perspektive auf den Werkbegriff kann man einnehmen, wenn man dem Ausgangspunkt von Steffen Martus' Studie zur *Werkpolitik* folgt, dass sich das Werk „um 1800“ als zentrale, auf einen Autor zugeordnete Instanz für die literarische Kommunikation³⁴ etabliert, um der Marktförmigkeit der literarischen Kommunikation gerecht zu werden. Über die Werkherrschaft des*r Autors*in, die diese*r durch den intentionalen Akt der Veröffentlichung und die Verfügungsgewalt über zusätzliches, hermeneutisch relevantes Material erlangt, stellt Martus seinen Begriff der ‚Werkpolitik‘, der die Herrschaftsverhältnisse über einen Text nicht unmittelbar dem Autor zuschreibt, sondern allgemeiner voraussetzt, „daß es überhaupt um Aushandlungen von Mächtigkeit geht, wenn sich Leser und Autoren mit einem Werk beschäftigen.“³⁵ Auf diese Weise rückt nicht nur das spezifische Autor-Text-Verhältnis zur Bestimmung eines Werks ins Zentrum, sondern eine Vielfalt institutioneller und kommunikativer Praxen. Dazu zählt Martus u. a. auch die Philologie, den Buchmarkt, das Bibliothekswesen, die Literaturkritik und die Leserschaft.

Bei solchen Analysen kann die Bestimmung von Werken in ihren para-, kon- oder intertextuellen Zusammenhängen helfen. Beispielsweise erwächst der ästhetische Wert der experimentellen Literatur vor allem aus der Qualität der ausgewählten Zitate sowie der Form der Montage.³⁶ Gérard Genette hat mit dem

³¹ Spoerhase (Anm. 27), 289.

³² Ebd., 288.

³³ Ebd., 289.

³⁴ Steffen Martus, *Werkpolitik. Zur Literaturgeschichte kritischer Kommunikation vom 17. bis zum 20. Jahrhundert*, Berlin/Boston 2007, 22.

³⁵ Ebd., 13.

³⁶ Vgl. Peter Bürger, *Theorie der Avantgarde*, Frankfurt a. M. ²1981 (1974), 97, 98, 107–111.

Begriff des *Paratexts* eine Kategorie eingeführt, die die von der Autorin (mit gestalteten Werkelemente neben dem ‚eigentlichen‘ Werk differenzierbar macht. Genette unterscheidet dabei *Peritexte*, also Elemente im direkten Umfeld des Werkes, wie Titel, Motto oder Danksagung, sowie *Epitexte*, das sind Elemente außerhalb des Werkes, zum Beispiel Autorinterviews zum Buch in anderen Medien oder die begleitende Verlagswerbung.³⁷ Annette Gilbert hat am Beispiel der Auseinandersetzungen über den Satz von Katharina Hackers *Alix, Anton und die anderen* (2009) zwischen der Autorin und dem *Suhrkamp Verlag* gezeigt, dass die „zunehmende (Verantwortungs-)Übernahme der Text- und Buchgestaltung durch die Autoren und die wachsende Ausdifferenzierung paratextueller Strategien [...] auch von philologischer und bibliothekarisch-archivarischer Seite Berücksichtigung finden“³⁸ müsse – wobei die Voraussetzung dieses Wandels die digitalen Textproduktionsmittel sind, die gleichsam zu den Autoren hin sozialisiert werden.

Eine weitere Problematisierung des Werkbegriffs ist mit der Kritik des literarischen Kanons verbunden: Der enge Bezug auf eine Autorintention und die formale Abgeschlossenheit eines literarischen Textes, die im statischen Werkbegriff angelegt sind, präferieren hochkulturelle ästhetische Autorschafts- und Literaturmodelle. Helmut Schmiedt hat eine *Krise des Werkbegriffs* konstatiert, indem er populärkulturelle Phänomene wie den Blues oder die Jugend- und Unterhaltungsliteratur kursorisch untersucht hat. Er sieht seine Hypothese bestätigt, dass sich „Gedanken zur Degradierung der Instanz ‚Autor‘ und zur Auflösung des Werkbegriffs [...] mit unseren Beispielen aus der U-Kultur trefflich und handfest belegen“ lassen.³⁹ Im Bereich der Jugend- und Unterhaltungsliteratur erreiche die Neube- und Überarbeitung der Texte ein solches Maß, „dass man vom ursprünglichen Text nur noch als von einer Vorlage für freie Improvisationen der Bearbeiter reden kann.“⁴⁰

Schließlich hat Roland Barthes paradigmatisch eine Bewegung *Vom Werk zum Text* eingefordert: Während das Werk zum Gegenstand der Philologie bzw. einer hermeneutischen Interpretation werde, versteht Barthes den Text strukturalistisch als „dezentriert, ohne Verriegelung“, als „ein System ohne Ende noch Zentrum“.⁴¹ Während das Werk ein Konsumgut sei, das eine klare Abstammung aufweise, löse der Text eine Lust bei den Leser*innen aus. Barthes ersetzt die enge Verbindung

³⁷Vgl. Gérard Genette, *Paratexte. Das Buch vom Beiwerk des Buches*, Frankfurt a. M. 2001 [frz. 1987], 9–21.

³⁸Annette Gilbert, „Unter ‚L‘ oder ‚F‘? Überlegungen zur Frage der Werkidentität bei literarischen Werken“, in: Dies. (Hg.), *Wiederaufgelegt. Zur Appropriation von Texten und Büchern in Büchern*, Berlin/Boston 2014, 67–85, hier: 83.

³⁹Helmut Schmiedt, „Tod des Autors und Krise des Werkbegriffs – als Phänomene der U-Kultur“, in: *KulturPoetik 3/2* (2003), 270–278, hier: 277.

⁴⁰Ebd., 273.

⁴¹Roland Barthes, „Vom Werk zum Text“, in: Stephan Kammer/Roger Lüdeke (Hg.), *Texte zur Theorie des Textes*, Stuttgart 2005, 40–51, hier: 44.

von Autor und Werk durch das lustvolle Verhältnis von Leser und Text: „[D]ie Gesellschaft postuliert eine Gesetzlichkeit des Bezugs zwischen dem Autor und seinem Werk [...]. Der ‚Text‘ hingegen wird ohne die Einschreibung des ‚Vaters‘ gelesen.“⁴² Barthes wählt, als nähme er das *World Wide Web* vorweg, als Metapher für den Text jene des ‚Netzes‘⁴³ und konstatiert, dass „der ‚Text‘ verlangt, daß man versucht, die Distanz zwischen Schreibenden und Lesen aufzuheben“.⁴⁴ Dieser später von Barthes – auch durch seine aphoristische Darstellungsform – noch radikalisierte Ansatz⁴⁵ wurde im literaturwissenschaftlichen Diskurs über den Werkbegriff eher verhalten aufgenommen und selbst in interessierten Auseinandersetzungen problematisiert.⁴⁶

3 Vom Werk zum Netzwerk? Netzliteratur und Literaturplattformen

Wenn man Barthes‘ Präferenz des Textbegriffs gegenüber dem Werkbegriff ernst nimmt und pragmatisch auf die konkrete Nutzung der Begriffe ‚Werk‘ und ‚Text‘ im deutschsprachigen Raum schaut sowie die für unsere Fragestellung relevanten Kategorien ‚Daten‘ und ‚Netzwerk‘ hinzunimmt, lässt sich vorsichtig ein Umschwung der Begriffsnutzung seit 1950 konstatieren. Folgt man den Daten des *Google Books Ngram Viewers* für den deutschsprachigen Korpus von 1720 bis 2008, kann für den Werkbegriff ein Aufstieg ab 1720, ein Höhepunkt seines Gebrauchs in den 1770er Jahren sowie eine konstante Etablierung mit einem deutlichen Anstieg von 1900 bis 1950 gezeigt werden, allerdings seit 1950 eine kontinuierliche Abnahme seiner Verwendung. Im Gegensatz dazu nimmt die Nutzung der Begriffe ‚Text‘ und vor allem (und seit etwa 1995 noch zunehmend) ‚Daten‘ deutlich zu. Während 1950 (in einer Hochphase der werkimmanenten Literaturanalyse) der Werkbegriff wesentlich häufiger genutzt wurde als der ‚Text‘- und der ‚Daten‘-Begriff, werden beide in der Gegenwart häufiger aufgerufen als der Werkbegriff. Seit den 1990er Jahren gewinnt zudem der Begriff des ‚Netzwerks‘ eine immer größere Relevanz (Abb. 1).

Diese Daten geben allerdings lediglich eine grobe Übersicht und erlauben nur vorsichtige Thesen. Man müsste auf Basis spezifischer Textkorpora aus verschiedenen Bereichen der Literaturwissenschaft die konkrete Verwendung dieser Begriffe analysieren. Simone Winko hat in einer empirischen Studie zur Anwendung des Werkbegriffs auf Texte von Heinrich von Kleist in der

⁴² Ebd., 46 f.

⁴³ Ebd., 47.

⁴⁴ Ebd., 48.

⁴⁵ Vgl. Roland Barthes, *Die Lust am Text*, Frankfurt a. M. 1974 [frz. 1973].

⁴⁶ Vgl. u. a. Margit Sutrop, „The Death of the Literary Work“, in: *Philosophy and Literature* 18/1 (1994), 38–49, hier: 48.



Abb. 1 Eine vergleichende Übersicht über die quantitative Nutzung der Begriffe ‚Daten‘, ‚Werk‘, ‚Text‘ und ‚Netzwerk‘ im deutschsprachigen Korpus von *Google Books* für die Jahre 1720 bis 2008⁴⁷

⁴⁷Nicht case-insensitive, with a smoothing of 5, vgl. *Google Books Ngram Viewer*, *Graph these comma-separated phrases: Daten, Netzwerk, Text, Werk*, https://books.google.com/ngrams/graph?content=Daten%2C+Netzwerk%2C+Text%2C+Werk&year_start=1720&year_end=2008&corpus=20&smoothing=5&share=&direct_url=t1%3B%2CDaten%3B%2C%3B.t1%3B%2CNetzwerk%3B%2C%3B.t1%3B%2CText%3B%2C%3B.t1%3B%2CWerk%3B%2C%3B.t1%3B%2C%3B (letzter Aufruf 22.6.2019).

literaturwissenschaftlichen Praxis in ähnlicher Weise konstatiert, dass der Werk- vom Textbegriff abgelöst worden sei und vor allem noch als Sammelbegriff für das *Œuvre* genutzt werde. Allerdings würden „Werk-Merkmale wie Intentionalität [...] auch auf die synonym verwendeten Begriffe wie ‚Text‘, editorische Bezeichnungen, den Autornamen oder Gattungsbezeichnungen übertragen.“⁴⁸

Für Einführungswerke in die Computerphilologie bzw. die *Digital Humanities* lässt sich vorsichtig eine Präferenz für andere Begriffe konstatieren: Fotis Jannidis nutzt den Werkbegriff nur im Kontext des gedruckten Ausgangstextes, ihm gegenüber stellt er ‚digitale Texte‘, ‚elektronische Texte‘, ‚elektronische Editionen‘ oder ‚Informationsbestände‘;⁴⁹ in einer anderen Einführung wird der Werkbegriff im Register nicht geführt, dort finden sich jedoch Begriffe wie ‚Daten‘, ‚Datensatz‘, ‚Digitalisat‘, ‚Textkorpora‘, ‚Volltext‘.⁵⁰ Jenseits dieser Korpusbeobachtungen, die einen Wandel vom Werkbegriff zu Konzeptionen von ‚Text‘ und ‚Daten‘ nahelegen, sollen im Folgenden grundsätzliche Überlegungen zum Verhältnis von Werkbegriffen und netzliterarischen Phänomenen angestellt werden.

3.1 Netzliteratur und Literaturplattformen: Organisation, Administration und Benutzungsschnittstellen

Die digitale Gesellschaft und ihre Medien werden als eine ‚Netzwerkgesellschaft‘ bzw. ‚Netzwerkultur‘ gefasst,⁵¹ der dezentralen Struktur des *World Wide Web* entsprechend. Während Jay David Bolter schon 1991 auf den ‚flüssigen‘ und dynamischen Charakter des elektronischen Schreibens hinweist und eine interaktive Computer-Literatur beschreibt,⁵² konstatieren zur Jahrtausendwende deutschsprachige Literaturwissenschaftler, dass die kommunikative Vernetzung und das kooperative Schreiben von Autoren und Lesern die *Literatur im elektronischen Raum* bzw. die *Netzliteratur* (mit)bestimmen werde.⁵³ Simone Winko grenzt dezidiert eine ‚digitalisierte (Buch-)Literatur‘ und eine ‚digitale

⁴⁸Winko nach: Weiß (Anm. 4), 423.

⁴⁹Vgl. Fotis Jannidis, „Computerphilologie“, in: Thomas Anz (Hg.), *Handbuch Literaturwissenschaft*, Bd. 2: *Methoden und Theorien*, Stuttgart/Weimar 2007, 27–40.

⁵⁰Fotis Jannidis/Hubertus Kohle/Malte Rehbein (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart 2017, 364–370.

⁵¹Vgl. u. a. Manuel Castells, *The Rise of the Network Society. The Information Age. Economy, Society and Culture*, Vol. 1, Oxford/Malden, MA, 1996; José van Dijck, *The Culture of Connectivity. A Critical History of Social Media*, Oxford 2013.

⁵²Jay D. Bolter, *Writing Space. The Computer, Hypertext, and the History of Writing*, Hillsdale, NJ, 1991, 4 u. 121–146.

⁵³Vgl. Christiane Heibach, *Literatur im elektronischen Raum*, Frankfurt a.M. 2003, 32–91; Peter Gendolla/Jörgen Schäfer, „Auf Spurensuche. Literatur im Netz, Netzliteratur und ihre Vorgeschichte(n)“, in: *Text+Kritik. Zeitschrift für Literatur*, H. 152: *Digitale Literatur*, Sonderband X (2001), 75–86, hier: 81.

Literatur‘ einerseits von einer „Netz-Literatur“ andererseits ab: Letztere bildet eine spezifische Untergattung der digitalen Literatur und benötigt das „Vermittlungsmedium“ Internet, „um zu funktionieren.“ Eine solche digitale Literatur, die die interaktiven Kommunikationsformen des Internets nutzt, kann „neue Möglichkeiten der Interaktion zwischen Text und Leser, aber auch der Kommunikation von Lesern miteinander“ hervorbringen.⁵⁴

Während Winko 2009 den Begriff der ‚Netzliteratur‘ in die Untergattungen ‚Hyperfictions‘, ‚Maschinenpoesie‘ und ‚Mitschreibprojekte‘ differenziert, haben sich durch die Wucherung der Sozialen Medien inzwischen vor allem intensiviertere netzliterarische Kommunikationsweisen, also ‚Mitschreibprojekte‘, entwickelt. Unter anderem lassen sich „Formen nutzergenerierter Texte“ in Medienformaten wie Weblogs, Microblogs, Wikis und Forenbeiträgen beschreiben,⁵⁵ wobei zwei verschiedene Kommunikationsweisen zu unterscheiden sind: einerseits Formen des digitalen kollaborativen Schreibens von Autor*innen und Leser*innen sowie andererseits *Social-Reading-Plattformen*, auf denen Leser*innen gemeinsam Texte rezipieren und kommentieren. In beiden Fällen erhalten die Leser*innen Zugang zu einem digitalen Text auf einer Literaturplattform, die es ihnen ermöglicht, den Leseakt als Schreibakt zu vollziehen (z. B. als Kommentar bei Weblogs, als eigenen Text in Foren oder als Teil eines Kollektivtexts in Wikis) und diesen Schreibakt – in unterschiedlicher Form – gleich wieder öffentlich zu machen (für alle Leser*innen oder für eine ausgewählte Gruppe).

Zwar werden die größten Umsätze auf dem elektronischen Literaturmarkt weiterhin mit Formaten gemacht, die als ‚E-Books‘ das gedruckte Buch und seine medialen Implikationen imitieren, zunehmend öffnet sich jedoch auch die Buchwissenschaft der Netzliteratur und ihren Erscheinungsweisen. Svenja Hagenhoff und Axel Kuhn verstehen digitale Lesemedien entsprechend nicht mehr als ‚Leseobjekte‘, sondern vielmehr als eine „dynamische Ausprägung[en] unterschiedlicher Eigenschaften“,⁵⁶ die nicht als eine modifizierte Form gedruckter Medien, sondern „als eigenständige Lesemedien mit spezifischer Kommunikationsfunktion“⁵⁷ zu betrachten seien. Aus einer solchen Perspektive wandelt sich das literarische Kommunikationsmodell von einer individuellen Buchautorschaft zur Autor-Leser-Interaktion in Sozialen Medien und von der Zentrierung des Werks als der zentralen (Verkaufs-)Größe des Buchmarkts hin zu literarischen Kommunikationsnetzwerken auf Webplattformen.

Dabei entstehen neue literarische Organisationsformen, wie Gesine Boesken für diese *Literaturplattformen* konstatiert, die auf „ein dezidiertes Plattform-„Management““ angewiesen sind, zu dem die jeweilige „Netiquette“, „spezifische

⁵⁴ Simone Winko, „Am Rande des Literaturbetriebs. Digitale Literatur im Internet“, in: Heinz Ludwig Arnold/Matthias Beilein (Hg.), *Literaturbetrieb in Deutschland*, München 2009, 292–303, hier: 294.

⁵⁵ Vgl. Kuhn/Kraus (Anm. 1), 684–688.

⁵⁶ Axel Kuhn/Svenja Hagenhoff, „Digitale Lesemedien“, in: Ursula Rautenberg/Ute Schneider (Hg.), *Lesen. Ein interdisziplinäres Handbuch*, Berlin/Boston 2015, 361–380, hier: 377.

⁵⁷ Ebd., 378.

Plattform-Regeln“ sowie die Etablierung von „Administratoren [...] und Moderatoren“ gehören.⁵⁸ Solche interaktiven Plattformen sind wiederum auf spezifische Benutzungsschnittstellen angewiesen, diese „übersetzen digital gespeicherten Code in maschinenlesbare Zeichen, gestalten Inhalte und zeigen diese an“.⁵⁹ Neben den eigentlichen literarischen Inhalt, der im Codex des gedruckten Buches eine festgelegte typographische Darstellung findet, tritt bei digitalen Lesemedien mit den Benutzungsschnittstellen eine Größe, die eine spezifische „Bereitstellungsqualität“ besitzen kann, wenn sie sich als „gestaltbar[]“ erweist.⁶⁰

Favorisiert der Begriff des ‚statischen Werks‘ gerade die (drucktechnische) Fixierung des literarischen Textes, ist die besondere Qualität der Benutzungsschnittstelle, dass sie eine vernetzte Kommunikation und ein Um- bzw. Weiter-schreiben des Textes ermöglicht. Die Geschäftsmodelle der Netzliteratur beziehen sich folglich nicht mehr auf die Produktion und Distribution eines gedruckten Buchwerks, sondern auf die Einbettung des literarischen Gegenstands in eine Netzkommunikation. Folglich experimentieren Autoren als *Self-Publisher*, aber auch Netzverlage sowie Anbieter von Literaturplattformen und Start-ups mit Geschäftsmodellen wie *Crowdfunding*, *Crowdsourcing*, *Flatrate*, *Freemium*, *Open Source*, *Pay-per-Use* oder *Pay What You Want*,⁶¹ die sich weniger auf die Werkförmigkeit des Gegenstands als eher auf seine Benutzungsschnittstellen und teilweise auch auf die Ermöglichung einer Anschlusskommunikation fokussieren.

3.2 *Fan-Fiction* und *Social Reading*: Werkbegriffe und juristische Probleme

Phänomene wie die Fan-Fiction, die Twitteratur, Litblogs mit Kommentarfunktion, literarische *Crowdfunding*- und/oder *Crowdsourcing*-Projekte oder auch die kollaborativen Annotationen zur Netzliteratur sowie die Formen der Online-Literaturkritik, z. B. auf Plattformen wie *Goodreads* oder *Mojoreads*, könnten Gegenstände genauerer Analysen sein.⁶² Dabei wäre es wichtig, die Gegenstände

⁵⁸ Gesine Boesken, „Literaturplattformen“, in: Christine Grond-Rigler/Wolfgang Straub (Hg.), *Literatur und Digitalisierung*, Berlin/Boston 2013, 21–42, hier: 29.

⁵⁹ Svenja Hagenhoff/Axel Kuhn, „Klickst Du noch oder liest Du schon? Softwarebasierte Benutzungsschnittstellen als Chance und Risiko digitaler Buch-, Zeitschriften- und Zeitungsangebote“, in: Sven Pagel (Hg.), *Schnittstellen (in) der Medienökonomie – Interaktion mit Medienpolitik, Medienrezeption und Medientechnologie*, Baden-Baden 2015, 217–239, hier: 217, <https://doi.org/10.5771/9783845264868-213>.

⁶⁰ Ebd., 236 u. 217.

⁶¹ Vgl. Oliver Gassmann/Karolin Frankenberger/Michaela Csik, *Geschäftsmodelle entwickeln. 55 innovative Konzepte mit dem St. Galler Business Model Navigator*, München 2013.

⁶² Vgl. die Analysen in Thomas Ernst, „Wem gehören Autor-Leser-Texte? Das geistige Eigentum, netzliterarische Standards, die Twitteratur von @tiny_tales und das Online-Schreibprojekt *morgen-mehr.de* von Tilman Rammstedt“, in: Sebastian Böck u. a. (Hg.), *Lesen X.0. Rezeptionsprozesse in der digitalen Gegenwart*, Göttingen 2017, 145–167, http://www.v-r.de/_uploads_media/files/9783847107453_boeck_etal_lesen_wz_082839.pdf (letzter Aufruf 1.5.2018).

nicht nur in ihrer Oberflächendarstellung, sondern auch in ihrer eigentlichen Form, dem jeweiligen *Code* wahrzunehmen; Martin Stobbe plädiert in diesem Sinne für den Erwerb einer „Procedural Literacy“.⁶³ Wenn man bspw. den Quellcode der Kommentare zu einem Blogposting auf dem Litblog *Die Dschungel. Anderswelt* von Alban Nikolai Herbst liest,⁶⁴ fällt es schwer, diesen im Kontext des Werks – also den Kommentar zum Posting im Kontext des Litblogs – zu rubrizieren: Die Kommentare könnten einerseits als eigenständige Werke mit einer Autorfunktion bewertet werden, andererseits gehören sie eindeutig zum Kommunikationszusammenhang des Litblogs, der als Medienformat auf diese Formen der Interaktion angelegt ist. Man könnte nun das Ausgangsblogposting als Werk und die Kommentare als fremdproduzierten *Peritext* bewerten; für die Fälle einer echten Autor-Leser-Kommunikation über die Kommentarfunktion erscheint eine solche Differenzierung jedoch ebenfalls unangemessen.

Zudem ergibt sich hier die auch juridische Frage, welche Elemente einer solchen Blogstruktur als Teile des Werks zu fassen wären: Während die literaturbetriebliche Werkzuschreibung an Bücher und Artikel über die Zuweisung einer ISBN oder einer ISSN erfolgt, die allerdings explizit nicht für Online-Datenbanken, Webseiten, E-Mails vergeben werden können, wären diese mit einem *Digital Object Identifier (DOI)* dauerhaft und eindeutig zu bezeichnen und aufzufinden.⁶⁵ Es ist allerdings weiterhin undeutlich, inwiefern die einzelnen Elemente der Autor-Leser-Kommunikation – z. B. im Kontext eines Litblogs – auch mit einem DOI zu versehen wären.

Diese grundsätzlichen kategorialen und urheberrechtlichen Probleme zeigen sich paradigmatisch bei der Fan-Fiction, die im deutschsprachigen Raum zum Beispiel auf der Plattform *FanFiktion.de* entsteht: Hier schreiben Leser*innen bzw. Zuschauer*innen ihre liebsten Bücher, Comics, Computerspiele, Filme und Fernsehserien weiter, indem sie sich eng an das Narrativ und/oder einzelne Figuren anlehnen. Die Fan-Fiction hat für eine große Schreibproduktivität vor allem junger Leser*innen gesorgt und bereits spezifische Merkmale ausgeprägt: eine hohe Intensität an gegenseitigen Kommentaren (als doppeltes *Social Reading*: erstens des Ausgangstexts, zweitens der Produktionen innerhalb der Community), spezifische Veröffentlichungsrhythmen und Rollen („Beta-Leser“ fungieren als Lektor*innen) sowie eigenständige Formate wie Mitmachprojekte oder Schreibwettbewerbe.⁶⁶

Allerdings ist die strukturelle Nähe der Fan-Fiction zu ihrer Vorlage urheberrechtlich bedenklich, weshalb die *Fan Fiction Studies* als ein zentrales Merkmal

⁶³Vgl. Martin Stobbe, „Quellcode lesen? Ein Plädoyer für Procedural Literacy in den Literaturwissenschaften“, in: Böck u. a. (Anm. 61), 47–68, hier: 64–66.

⁶⁴Vgl. Alban N. Herbst (2017), <http://albannikolaiherbst.twoday.net/stories/als-huttraeger/#comments> (letzter Aufruf 1.5.2018).

⁶⁵Vgl. Ralph O. Graef, *Recht der E-Books und des Electronic Publishing*, München 2016, 157 u. 159.

⁶⁶Vgl. Wolfgang Reißmann/Nadine Klass/Dagmar Hoffmann, „Fan Fiction, Urheberrecht und Empirical Legal Studies“, in: *Pop. Kultur & Kritik* 10 (2017), 156–172, hier: 160.

der Fan-Fiction den Willen, offensiv das Recht zur kreativen Nachnutzung zu proklamieren, definieren: „[F]ans are self-consciously organizing and establishing their rights to create transformative works.“⁶⁷ Dennoch sind die aktiven Leser-Autoren der Bedrohung ausgesetzt, dass ihre Fan-Fiction als Plagiat bewertet und sie entsprechend bestraft werden können. Reißmann u. a. schlagen – im Kontext ihrer *Empirical Legal Studies* zur Fan-Fiction – vor, gleich mehrere Annahmen des deutschen Urheberrechts kritisch zu überprüfen, die sich auf den juristischen Werkbegriff und seine Bedeutung für das Urheberrecht beziehen (ich führe die ersten vier von sechs Vorschlägen auf): Überprüft werden sollte

- die Fixierung auf den Individualschöpfer;
- das vermutete starke ‚persönliche Band‘ zwischen dem Schöpfer und seinem Werk;
- das von einer rigiden Bestimmung von Selbstständigkeit geprägte Werkverständnis;
- der mit Blick auf Werkarten und Schaffensprozesse favorisierte *one-size-fits-all*-Ansatz des deutschen Urheberrechtsgesetzes [...].⁶⁸

Zumindest die ersten beiden Punkte interferieren direkt mit dem im literarischen und im literaturwissenschaftlichen Diskurs präsenten statischen Werkbegriff. Die Fan-Fiction als eine produktive Praxis, die den Leseakt zu einem Schreibakt macht, steht repräsentativ für die Schwierigkeiten, die der statische und auf die Verbindung von Autor und Werk fixierte Werkbegriff für eine digitale literarische Kommunikation mit sich bringt, in der Autor*innen und Leser*innen sowie Leser*innen und Leser*innen enger zusammenrücken und die sich in öffentlichen Formen einer kollaborativen Textproduktion niederschlägt.

Im rechtswissenschaftlichen Diskurs werden diese Probleme zunehmend angegangen. Claudia Summerer empfiehlt in ihrer Arbeit ‚*Illegale Fans*‘ eine modifizierte Auslegung des § 24 UrhG, der garantiert, dass ein „selbständiges Werk, das in freier Benutzung des Werkes eines anderen geschaffen worden ist, [...] ohne Zustimmung des Urhebers des benutzten Werkes veröffentlicht und verwertet werden“⁶⁹ kann. Da es im Fall der Fan-Fiction jedoch uneindeutig ist, inwiefern die Fan-Texte als ‚selbstständige Werke‘ zu fassen sind, schlägt Summerer vor, eine allgemeiner gehaltene „Ausnahmeregelung für kreative Werknutzungen“⁷⁰ zu ergänzen.

⁶⁷ Karen Hellekson/Kristina Bosse, „Introduction. Why a Fan Fiction Studies Reader Now?“, in: Dies. (Hg.), *The Fan Fiction Studies Reader*, Iowa City 2014, 1–17, hier: 16.

⁶⁸ Reißmann/Klass/Hoffmann (Anm. 65), 164.

⁶⁹ Bundesministerium der Justiz und für Verbraucherschutz (Anm. 25), § 24.

⁷⁰ Claudia Summerer, ‚*Illegale Fans*‘. *Die urheberrechtliche Zulässigkeit von Fan Art*, Berlin/Boston 2015, 199.

Während sich viele rechtswissenschaftliche Beiträge zwar vor allem mit urheberrechtlichen Schutz-, Sanktions- und Haftungsmaßnahmen im Kontext digitaler Medien⁷¹ und dem *Digital Rights Management* als einem Lösungsversuch⁷² befassen, bemühen sich andere um eine differenzierte Auseinandersetzung mit den neuen Medienformaten und stellen dabei auch die zentrale Größe des Urheberrechts, wie die Vorstellung des komplexen Werks ‚Computerspiel‘ als einer individuellen Schöpfung, infrage.⁷³ Zudem zeichnet sich eine Tendenz ab, die „Unzulänglichkeiten des urheberzentrierten Individualitätskonzepts“⁷⁴ hervorzuheben: Malek Barudi zeigt in *Autor und Werk*, dass die vom Urheberrecht zentral gesetzte ‚Prägetheorie‘, der zufolge sich in einem literarischen Werk die individuelle Autorschaft erweise, nicht angemessen überprüfbar sei. Anstelle dessen plädiert er für den Bezug auf das Werk selbst und seine Abgrenzung von anderen Werken.⁷⁵

Für den spezifischen Bereich der Netzliteratur sind auch die Arbeiten von Axel Metzger relevant, der das Urheberrecht und die *Freie oder Open Source Software (FOSS)* miteinander in ein Verhältnis setzt. Eine Vergleichbarkeit der Phänomene ist gegeben, da beide auf die kollaborative Produktion sowie die offene Verfügbar- und Nachnutzbarkeit angelegt sind. Schon 2010 konstatierte Metzger eine starke Differenz zwischen „dem im Gesetz mit Nachdruck betonten Schöpferprinzip und der weniger prinzipientreuen Rechtsrealität“,⁷⁶ insbesondere für „dezentrale Urheberkollektive“⁷⁷ sei das Urheberrecht wenig hilfreich. In einem späteren Beitrag untersucht er diese Frage genauer am Beispiel der *Wikimedia Foundation* und stellt fest, dass die (offene) Lizenzierung von Inhalten die führende Tendenz sei. In solchen Schreibnetzwerken seien allerdings die unterschiedlichen

⁷¹ Vgl. u. a. Thomas Hoeren/Viola Bensinger (Hg.), *Haftung im Internet. Die neue Rechtslage*, Berlin/Boston 2014; Chiara Santangelo, *Der urheberrechtliche Schutz digitaler Werke. Eine vergleichende Untersuchung der Schutz- und Sanktionsmaßnahmen im deutschen, italienischen und englischen Recht*, Berlin 2011; Katharina Ziegler, *Urheberrechtsverletzungen durch Social Sharing. Urheber- und haftungsrechtliche Aspekte sozialer Netzwerke am Beispiel der Plattform Facebook*, Tübingen 2016.

⁷² Vgl. u. a. Gunda Dreyer, „Urheberrechtliche Problembereiche des Digital Rights Managements“, in: Louis Pahlow/Jens Eisfeld (Hg.), *Grundlagen und Grundfragen des Geistigen Eigentums*, Tübingen 2008, 221–250.

⁷³ Vgl. Claas Oehler, *Komplexe Werke im System des Urheberrechtsgesetzes am Beispiel von Computerspielen. Zugleich ein Beitrag zur Auslegung der §§ 8 und 9 UrhG*, Baden-Baden 2016, 485.

⁷⁴ Barudi (Anm. 21), 199.

⁷⁵ Ebd., 200.

⁷⁶ Axel Metzger, „Vom Einzelurheber zu Teams und Netzwerken. Erosion des Schöpferprinzips?“, in: Stefan Leible/Ansgar Ohly/Herbert Zech (Hg.), *Wissen – Märkte – Geistiges Eigentum*, Tübingen 2010, 79–92, hier: 82.

⁷⁷ Ebd., 92.

Autorfunktionen, Schreibrechte, „Streitbeilegungsmechanismen“, „Community-Regularien“ sowie deren „Auswirkungen [...] auf die kreativen Prozesse“⁷⁸ zentral.

4 Literarische Netzwerkpolitik. Fazit und Ausblick

Der Werkbegriff als eine zentrale Kategorie der Literaturwissenschaft ist historisch und medial starken Wandlungen unterworfen, die jeweils die Ökonomie der literarischen Kommunikation – zwischen Autor*in, Werk/Text, den Instanzen und Institutionen des Literaturbetriebs und den Leser*innen – in unterschiedlicher Weise strukturiert haben. Zwar hat sich gezeigt, dass der Gebrauch des Werkbegriffs seit 1950 zurückgeht (und an seine Stelle Begriffe wie ‚Text‘ und ‚Daten‘ treten), seine Gehalte jedoch noch immer wirksam sind. Dazu konnte modellhaft ein statischer von einem dynamischen Werkbegriff unterschieden werden: Während im statischen Werkbegriff ein literarisches Werk vor allem auf die Autorintention bezogen und in der Form des fixierten Druckwerks als enthoben gedacht wird, beschreibt der dynamische Werkbegriff breitere und pragmatische literarische Kommunikationsmodelle: Das Werk steht zwischen Autor und Leser als Gegenstand einer Werkpolitik, in die auch andere Akteure eingreifen, und im Zusammenhang von Para-, Inter- und Kontexten.

Die Netzliteratur erscheint als ein vernetztes Soziales Medium, ihre neuen Formate wie die auf Literaturplattformen produzierten Fan-Fiction oder Litblogs intensivieren die Autor-Leser-Kommunikation und produzieren Textfragmente, deren Status noch erforscht wird.⁷⁹ Für eine Netzliteraturwissenschaft, die sich als Teildisziplin der *Digitalen Literaturwissenschaft* um die Analyse dieser netzliterarischen Phänomene bemüht, ist eine Untersuchung der Literaturplattformen und ihrer Benutzerschnittstellen unabdingbar: Wie sind diese gestaltet, welche (Mit-)Schreibrechte wurden an welche Rollen vergeben, welche Interaktivitätsoptionen bestehen, wie wird die Netzkommunikation reguliert? Die Perspektive wird verschoben: Der literarische Gegenstand ist weniger das mit einem Placet des Autors versehene fixierte Produkt eines Schreibprozesses, sondern ein literarisches Kommunikationsnetzwerk, das (teilweise) seinen Produktionsprozess offenlegt und insbesondere seinen Rezeptionsprozess als einen des Mitschreibens anlegt – und gerade durch die Schnittstellennutzbarkeit seine Marktfähigkeit beweist.

⁷⁸Axel Metzger, „Open Source und andere alternative Lizenzmodelle. Nutzung des Urheberrechts als Mittel zur Sicherung freier Kommunikation“, in: Michael Grünberger/Stefan Leible (Hg.), *Die Kollision von Urheberrecht und Nutzerverhalten*, Tübingen 2014, 115–129, hier: 127.

⁷⁹In der in der *Villa Vigoni* diskutierten ersten Beitragsversion habe ich dafür plädiert, diese spezifische Form der Autor-Leser-Kommunikation unter Roland Barthes' Textbegriff zu fassen. Durch die Diskussionen in der *Villa Vigoni* sehe ich dies kritisch und plädiere nun wesentlich stärker für eine Nutzung des Netzwerkbegriffs und für seine kritische Reflexion in der *Netzliteraturwissenschaft* als einer Teildisziplin der *Digitalen Literaturwissenschaft*.

Diese Verschiebung vergrößert jedoch zugleich die Legitimationskrise der durch die urheberrechtliche Prägetheorie legitimierten Rechts- und Geschäftsmodelle des literarischen Publizierens. Für die literaturwissenschaftliche Arbeit wird es folglich weniger relevant, „Erkenntnisse aus der Forschung zu analogen Printmedien unreflektiert auf digitale Lesemedien übertragen“ und in eine Relation zu Größen wie „Text/Hypertext, Buch/E-Book oder Papier/Bildschirm“⁸⁰ zu setzen. Vielmehr sollte den spezifischen digitalen Bedingungen der netzliterarischen Produktion und Rezeption nachgegangen werden. Im Zentrum muss dabei die Analyse der *Netzwerkpolitik* stehen, also der literarischen Online-Kommunikationsnetze, auf Basis eines interdisziplinären Wissenstransfers zwischen Literatur-, Medien-, Buch- und Rechtswissenschaft.

Ergänzung 2021: Vom Werk zum Netzwerk? Differenzierung vs. Restauration vs. Überwindung des literaturwissenschaftlichen Werkbegriffs

Seit 2017 hat sich eine ‚Netzliteraturwissenschaft‘ noch intensiver formiert, die als Teil der Digitalen Literaturwissenschaft auf die Herausforderung reagiert, Literatur und literarische Kommunikation in Sozialen Medien bzw. im World Wide Web zu erforschen.⁸¹ Ein wichtiger netzliteraturwissenschaftlicher Forschungsbereich beschäftigt sich mit der Frage, inwiefern sich durch die veränderten medialen und kommunikativen Bedingungen auch die Bedeutungen und Beziehungen zentraler literaturwissenschaftlicher Kategorien wie ‚Autorschaft‘, ‚Lesen‘ oder ‚Werk‘ wandeln.

Die Beziehungen zwischen kulturellen Produzent*innen und Leser*innen, der Status der künstlerischen ‚Werke‘ und die produktiven Rezeptionsprozesse im digitalen Raum müssen neu austariert werden. Wie dynamisch und gesellschaftlich brisant diese Kämpfe sind, zeigte sich 2019 in den Debatten über die neuen Urheberrechtsrichtlinien auf der Ebene der Europäischen Union und ihrer Mitgliedsstaaten, insbesondere in den Kontroversen um das ‚Leistungsschutzrecht‘ und die ‚Uploadfilter‘. Vor diesen Hintergründen erscheint es unabdingbar, dass der durch die Digitalisierung veränderte Status literarischer Werkformen auch literaturwissenschaftliche Forschungsarbeiten hervorbringt, die entweder 1) eine Relektüre und Differenzierung des Werkbegriffs und seiner Geschichte intendieren oder aber 2) ihn explizit zu stärken versuchen oder aber 3) ihn in einer Auseinandersetzung mit netzliterarischen Phänomenen problematisieren bzw. neue Konzeptionen oder modifizierte Begrifflichkeiten vorschlagen.

⁸⁰ Hagenhoff/Kuhn (Anm. 58), 222.

⁸¹ Vgl. Netzliteraturwissenschaft – Forschung und Community (2021) <https://netzliteraturwissenschaft.net/> (letzter Aufruf 10.09.2021). Die Notwendigkeit, ein solches Forschungsfeld zu systematisieren, ist 2017 durch die Konferenz „Digitale Literaturwissenschaft“, deren Ergebnisse dieser Band abbildet, in besonderer Weise offensichtlich geworden, vgl. Thomas Ernst, *Netzliteraturwissenschaft. Begründung und Systematisierung einer Disziplin*, Berlin/Boston 2022 [in Arbeit].

Mit seiner Münsteraner Dissertationsschrift legt Thomas Kater eine Arbeit vor, die das ‚Werk‘ als Status und als Spezifikation von ‚Text‘ reflektiert und Grundlinien eines pragmatischen Werkbegriffs entwickelt, den er in ‚Werkstatusfunktionen‘ differenziert. Am Beispiel des Œuvres von Max Frisch untersucht Kater ästhetische, epistemische, moralische und rechtliche, politische und ökonomische Funktionen des Werkstatus.⁸² Annette Gilbert befasst sich in ihrer Habilitationsschrift mit „Grenzfälle[n] literarischer Werkwerdung seit den 1950er Jahren“. Gilbert problematisiert Kategorien des Werkstatus, der Werkmaterialität und -integrität, der Werkzuschreibung und -identität und der Werksozialität und -autonomie am Beispiel von zahlreichen literarischen Gegenständen, die allesamt die Grenzen der Kategorie ‚Werk‘ herausfordern.

Auch der 2019 von Lutz Danneberg, Annette Gilbert und Carlos Spoerhase herausgegebene Band „Das Werk. Zum Verschwinden und Fortwirken eines Grundbegriffs“ stützt den Werkbegriff als eine zentrale Kategorie literaturwissenschaftlicher Arbeit mit zahlreichen differenziert argumentierenden Aufsätzen. Allerdings sei hier auch auf die Position von Werner Wolf verwiesen, der in besonderer Weise für die „Sinnhaftigkeit der Restauration bzw. Wiederverwendung des Werkbegriffs als eines Grundkonzeptes nicht nur der Literaturwissenschaft“ plädiert. Dazu greift er u. a. auf Vorstellungen der Einmaligkeit großer Kunstwerke, der künstlerischen Intention und der wichtigen Abgrenzung einer Hoch- von der Populärkultur zurück.⁸³

In eine andere Richtung weisen Forschungsarbeiten, die sich explizit um die Problematisierung des Werkbegriffs im Kontext der Netzliteratur oder des digitalen Publizierens bemühen. In der von Svetlana Efimova herausgegebenen *Textpraxis*-Sonderausgabe „Autor und Werk. Wechselwirkungen und Perspektiven“ (2018) sticht der Beitrag von Julia Nantke hervor, der am Beispiel der digitalen bzw. vernetzten Literatur nicht nur über kooperative Projekte und das Werk als Dialog nachdenkt, sondern auch über die (un-)begrenzten Anschlusskommunikationen in Mitschreibprojekten, die Rolle der aktiven Leser*innen „zwischen Kommentatoren und Kollaborateuren“, über „Autorenkollektive“ und „Mensch-Maschine-Kollaborationen“ sowie über das Verhältnis von multiplen Autorschaften und digitalen Werken: „[D]as Paradigma multipler Autorschaft

⁸² Kater, Thomas, *Status und Funktion. Zu Theorie und Praxis des literarischen Werks. Mit Analysen zum Œuvre von Max Frisch*, Münster 2021 [unveröff. Diss.].

⁸³ Vgl. Lutz Danneberg/Annette Gilbert/Carlos Spoerhase (Hg.), *Das Werk. Zum Verschwinden und Fortwirken eines Grundbegriffs*, Berlin/Boston 2019; Werner Wolf, „Du texte à l'oeuvre? Zur Sinnhaftigkeit der Restauration bzw. Wiederverwendung des Werkbegriffs als eines Grundkonzeptes nicht nur der Literaturwissenschaft“, in: Lutz Danneberg/Annette Gilbert/Carlos Spoerhase (Hg.), *Das Werk. Zum Verschwinden und Fortwirken eines Grundbegriffs*, Berlin/Boston 2019, 379–396.

und daran geknüpft des dynamisierten Werks ermöglichen einen neuen Blick auf literarische Produktions- und Publikationsprozesse, die verstärkt in ihrer Netzwerkhaftigkeit perspektiviert werden“.⁸⁴

Schließlich zeigt das Working Paper „Digitales Publizieren in den Geisteswissenschaften: Begriffe, Standards, Empfehlungen“⁸⁵, das die AG Digitales Publizieren im DHd-Verband 2021 im Open Peer Review veröffentlicht hat, wie stark die veränderten Publikationspraktiken der Digital Humanities auch die Begrifflichkeiten verändern: In dieser Veröffentlichung, zu der auch Literaturwissenschaftler*innen beigetragen haben, wird der Werkbegriff nur noch zweimal genutzt (beide Male als ‚gedrucktes Werk‘ in Abgrenzung zum ‚digitalen‘ oder ‚elektronischen Text‘), während andere Konzeptualisierungen wie ‚Text‘ oder ‚Version‘ etwa zwanzig- bzw. dreißigmal herangezogen werden.

Literatur⁸⁶

- AG Digitales Publizieren, „*Digitales Publizieren in den Geisteswissenschaften. Begriffe, Standards, Empfehlungen*“ (2021), https://doi.org/10.17175/wp_2021_001.
- Barthes, Roland, *Die Lust am Text*, Frankfurt a.M. 1974 [frz. 1973].
- Barthes, Roland, „Vom Werk zum Text“, in: Stephan Kammer/Roger Lüdeke (Hg.), *Texte zur Theorie des Textes*, Stuttgart 2005, 40–51.
- Barudi, Malek, *Autor und Werk – eine prägende Beziehung. Die urheberrechtliche Prägetheorie im Spiegel der Literaturwissenschaft*, Tübingen 2013.
- Boesken, Gesine, „Literaturplattformen“, in: Christine Grond-Rigler/Wolfgang Straub (Hg.), *Literatur und Digitalisierung*, Berlin/Boston 2013, 21–42.
- Bolter, Jay D., *Writing Space. The Computer, Hypertext, and the History of Writing*, Hillsdale, NJ, 1991.
- Bosse, Heinrich, *Autorschaft ist Werkherrschaft. Über die Entstehung des Urheberrechts aus dem Geist der Goethezeit*, Paderborn u.a. 1981.
- Brockhaus, *Brockhaus Enzyklopädie*, Band 29: VERTI-WETY, Mannheim 2006.
- Bürger, Peter, *Theorie der Avantgarde*, Frankfurt a.M. 1981 (1974).
- Castells, Manuel, *The Rise of the Network Society. The Information Age. Economy, Society and Culture*, Vol. 1, Oxford/Malden, MA, 1996.
- Conrad, Maren/Schmidtke, Theresa/Stobbe, Martin (Hg.), *Textpraxis. Digitales Journal für Philologie* 8/2 (2017), Sonderausgabe #2: *Digitale Kontexte. Literatur und Computerspiel in der Gesellschaft der Gegenwart*. <https://www.textpraxis.net/node/332>.

⁸⁴ Julia Nantke, „Multiple Autorschaft als digitales Paradigma und dessen Auswirkungen auf den Werkbegriff“, in: Svetlana Efimova (Hg.), *Autor und Werk. Wechselwirkungen und Perspektiven*. Sonderausgabe # 3 von *Textpraxis. Digitales Journal für Philologie* 2 (2018), <http://www.textpraxis.net/julia-nantke-multiple-autorschaft>, <http://dx.doi.org/10.17879/77159516645> (letzter Aufruf 25.10.2021); Svetlana Efimova (Hg.), *Autor und Werk. Wechselwirkungen und Perspektiven*. Sonderausgabe # 3 von *Textpraxis. Digitales Journal für Philologie* 2 (2018), <https://www.textpraxis.net/sonderausgabe-3> (letzter Aufruf 25.10.2021).

⁸⁵ AG Digitales Publizieren, „*Digitales Publizieren in den Geisteswissenschaften. Begriffe, Standards, Empfehlungen*“ (2021), https://doi.org/10.17175/wp_2021_001.

⁸⁶ Sämtliche Online-Referenzen wurden letztmalig am 1.5.2018 eingesehen. Die Referenzen der Fußnoten 80ff. wurden letztmalig am 25.10.2021 eingesehen.

- Danneberg, Lutz/Gilbert, Annette/Spoerhase, Carlos (Hg.), *Das Werk. Zum Verschwinden und Fortwirken eines Grundbegriffs*, Berlin/Boston 2019.
- Dijk, José van, *The Culture of Connectivity. A Critical History of Social Media*, Oxford 2013.
- Dommann, Monika, *Autoren und Apparate. Die Geschichte des Copyrights im Medienwandel*, Frankfurt a.M. 2014.
- Dreyer, Gunda, „Urheberrechtliche Problembereiche des Digital Rights Managements“, in: Louis Pahlow/Jens Eisfeld (Hg.), *Grundlagen und Grundfragen des Geistigen Eigentums*, Tübingen 2008, 221–250.
- Efimova, Svetlana (Hg.), *Autor und Werk. Wechselwirkungen und Perspektiven*. Sonderausgabe # 3 von *Textpraxis. Digitales Journal für Philologie* 2 (2018), <https://www.textpraxis.net/sonderausgabe-3> (letzter Aufruf 25.10.2021).
- Ernst, Thomas, *Literatur als soziales Medium in der digitalisierten Gesellschaft. Zur Begründung einer Netzliteraturwissenschaft*, Essen 2018 (unveröff. Habilitationsschrift).
- Ernst, Thomas, „Wem gehören Autor-Leser-Texte? Das geistige Eigentum, netzliterarische Standards, die Twitteratur von @tiny_tales und das Online-Schreibprojekt *morgen-mehr.de* von Tilman Rammstedt“, in: Sebastian Böck u.a. (Hg.), *Lesen X.0. Rezeptionsprozesse in der digitalen Gegenwart*, Göttingen 2017, 145–167, http://www.v-r.de/_uploads_media/files/9783847107453_boeck_etal_lesen_wz_082839.pdf.
- Ernst, Thomas, *Netzliteraturwissenschaft. Begründung und Systematisierung einer Disziplin*, Berlin/Boston 2022 [in Arbeit].
- Foucault, Michel, „Was ist ein Autor?“, in: Ders.: *Schriften zur Literatur*, aus dem Französischen von Karin von Hofer und Anneliese Botond, Frankfurt a.M. 1988, 7–31 [frz. 1969].
- Foucault, Michel, *Die Ordnung des Diskurses*, mit einem Essay von Ralf Konersmann, Frankfurt a.M. 2003 [frz. 1972].
- Gassmann, Oliver/Frankenberger, Karolin/Csik, Michaela, *Geschäftsmodelle entwickeln. 55 innovative Konzepte mit dem St. Galler Business Model Navigator*, München 2013.
- Gendolla, Peter/Schäfer, Jörgen, „Auf Spurensuche. Literatur im Netz, Netzliteratur und ihre Vorgeschichte(n)“, in: *Text + Kritik. Zeitschrift für Literatur*. H. 152: *Digitale Literatur*, Sonderband X (2001), 75–86.
- Genette, Gérard, *Palimpseste. Die Literatur auf zweiter Stufe*, Frankfurt a.M. 1993 [frz. 1982].
- Genette, Gérard, *Paratexte. Das Buch vom Beiwerk des Buches*, Frankfurt a.M. 2001 [frz. 1987].
- Gerstenberg, Ekkehard, *Die Urheberrechte an Werken der Architektur und der Photographie*, München 1968.
- Gilbert, Annette, „Unter ‚L‘ oder ‚F‘? Überlegungen zur Frage der Werkidentität bei literarischen Werken“, in: Dies. (Hg.), *Wiederaufgelegt. Zur Appropriation von Texten und Büchern in Büchern*, Berlin/Boston 2014, 67–85.
- Gilbert, Annette, *Im toten Winkel der Literatur. Grenzfälle literarischer Werkwerdung seit den 1950er Jahren*, Paderborn 2018.
- Graef, Ralph O., *Recht der E-Books und des Electronic Publishing*, München 2016.
- Hagenhoff, Svenja/Kuhn, Axel, „Klickst Du noch oder liest Du schon? Softwarebasierte Benutzungsschnittstellen als Chance und Risiko digitaler Buch-, Zeitschriften- und Zeitungsangebote“, in: Sven Pagel (Hg.), *Schnittstellen (in) der Medienökonomie – Interaktion mit Medienpolitik, Medienrezeption und Medientechnologie*, Baden-Baden 2015, 217–239, DOI <https://doi.org/10.5771/9783845264868-213>.
- Hahn, Richard/Schippa, Martin, *Rechtsfragen im Verlag. Urheberrecht, Verlagsrecht & Co*, Berlin/Boston 2014.
- Heibach, Christiane, *Literatur im elektronischen Raum*, Frankfurt a.M. 2003.
- Hellekson, Karen/Bosse, Kristina, „Introduction. Why a Fan Fiction Studies Reader Now?“, in: Dies. (Hg.), *The Fan Fiction Studies Reader*, Iowa City 2014, 1–17.
- Hoeren, Thomas/Bensinger, Viola (Hg.), *Haftung im Internet. Die neue Rechtslage*, Berlin/Boston 2014.
- Iser, Wolfgang, *Die Appellstruktur der Texte. Unbestimmtheit als Wirkungsbedingung literarischer Prosa*, Konstanz 1970.

- Jannidis, Fotis, „Computerphilologie“, in: Thomas Anz (Hg.), *Handbuch Literaturwissenschaft*, Bd. 2: *Methoden und Theorien*, Stuttgart/Weimar 2007, 27–40.
- Jannidis, Fotis/Kohle, Hubertus/Rehbein, Malte (Hg.), *Digital Humanities. Eine Einführung*, Stuttgart 2017.
- Kater, Thomas, „Keine Wiederkehr des Dagewesenen. Zu Gegenwart und Potential des literarischen Werkbegriffs“ (Conference Proceedings of: Wiederkehr des Werks? Symposium zur Gegenwart des literarischen Werkbegriffs. Schloss Herrenhausen, Hannover, 21.-23. Oktober 2015.), in: *JLToonline – Journal of Literary Theory* (8.3.2016), <http://www.jltonline.de/index.php/conferences/article/view/783/1838>, URN nbn:de:0222-003213.
- Kater, Thomas, *Status und Funktion. Zu Theorie und Praxis des literarischen Werks. Mit Analysen zum Œuvre von Max Frisch*, Münster 2021 [unveröff. Diss.].
- Köbel, Martin, „Das literarische Werk. Zur Geschichte eines Grundbegriffs der Literaturtheorie“, in: *Text.Kritische Beiträge* 10 (2005), 27–44.
- Kreknin, Innokentij/Marquardt, Chantal (Hg.), *Textpraxis. Digitales Journal für Philologie* 7/2 (2016), Sonderausgabe #1: *Das digitalisierte Subjekt. Grenzbereiche zwischen Fiktion und Alltagswirklichkeit*.
- Kristeva, Julia, „Bachtin, das Wort, der Dialog und der Roman“, in: Jens Ihwe (Hg.), *Literaturwissenschaft und Linguistik. Ergebnisse und Perspektiven, Bd. 3: Zur linguistischen Basis der Literaturwissenschaft II*, Frankfurt a.M. 1972, 345–375 [frz. 1967].
- Kuhn, Axel/Hagenhoff, Svenja, „Digitale Lesemedien“, in: Ursula Rautenberg/Ute Schneider (Hg.), *Lesen. Ein interdisziplinäres Handbuch*, Berlin/Boston 2015, 361–380.
- Kuhn, Axel/Kraus, Susanne, „Nutzergenerierte Texte in digitalen Netzwerken“, in: Ursula Rautenberg/Ute Schneider (Hg.), *Lesen. Ein interdisziplinäres Handbuch*, Berlin/Boston 2015, 679–699.
- Link, Jürgen/Link-Heer, Ursula, „Diskurs/Interdiskurs und Literaturanalyse“, in: *Zeitschrift für Literaturwissenschaft und Linguistik* 20 (1990), 88–99.
- Link, Jürgen/Parr, Rolf, „Semiotik und Interdiskursanalyse“, in: Klaus-Michael Bogdal (Hg.), *Neue Literaturtheorien. Eine Einführung*, Göttingen 2005, 108–133.
- Martens, Gunther, „Das Werk als Grenze. Ein Versuch zur terminologischen Bestimmung eines editorischen Begriffs“, in: *editio* 18 (2004), 175–186.
- Martus, Steffen, *Werkpolitik. Zur Literaturgeschichte kritischer Kommunikation vom 17. bis zum 20. Jahrhundert*, Berlin/Boston 2007.
- Metzger, Axel, „Vom Einzelurheber zu Teams und Netzwerken. Erosion des Schöpferprinzips?“, in: Stefan Leible/Ansgar Ohly/Herbert Zech (Hg.), *Wissen – Märkte – Geistiges Eigentum*, Tübingen 2010, 79–92.
- Metzger, Axel, „Open Source und andere alternative Lizenzmodelle. Nutzung des Urheberrechts als Mittel zur Sicherung freier Kommunikation“, in: Michael Grünberger/Stefan Leible (Hg.), *Die Kollision von Urheberrecht und Nutzerverhalten*, Tübingen 2014, 115–129.
- Müller, Harro, „Einige Notizen zu Diskurstheorie und Werkbegriff“, in: Jürgen Fohrmann/Harro Müller (Hg.), *Diskurstheorien und Literaturwissenschaft*, Frankfurt a.M. 1988, 235–243.
- Nantke, Julia, „Multiple Autorschaft als digitales Paradigma und dessen Auswirkungen auf den Werkbegriff“, in: Svetlana Efimova (Hg.), *Autor und Werk. Wechselwirkungen und Perspektiven*. Sonderausgabe # 3 von *Textpraxis. Digitales Journal für Philologie* 2 (2018), <http://www.textpraxis.net/julia-nantke-multiple-autorschaft>, <https://doi.org/10.17879/77159516645> (letzter Aufruf 25.10.2021).
- Oehler, Claas, *Komplexe Werke im System des Urheberrechtsgesetzes am Beispiel von Computerspielen. Zugleich ein Beitrag zur Auslegung der §§ 8 und 9 UrhG*, Baden-Baden 2016.
- Ostermann, Eberhard, „Werk“, in: Gert Ueding (Hg.), *Historisches Wörterbuch der Rhetorik*, Band 9: *St-Z*, Tübingen 2009, 1352–1362.
- Parr, Rolf, „Interdiskurstheorie/Interdiskursanalyse“, in: Ders./ Clemens Kammler/Ulrich J. Schneider (Hg.), *Foucault-Handbuch. Leben, Werk, Wirkung*, Stuttgart 2008, 202–206.
- Platon, „Phaidros“, in: Ders., *Sämtliche Werke*, hg. von Erich Loewenthal. Bd. 2, Berlin/Heidelberg 1940 (gr. um 370 v.u.Z.), 411–482, <http://www.zeno.org/nid/20009262660>.

- Reißmann, Wolfgang/Klass, Nadine/Hoffmann, Dagmar, „Fan Fiction, Urheberrecht und Empirical Legal Studies“, in: *Pop. Kultur & Kritik* 10 (2017), 156–172.
- Rockenberger, Annika/Röcken, Per, „Wie ‚bedeutet‘ ein ‚material text‘?“, in: Wolfgang Lukas/Rüdiger Nutt-Kofoth/Madleen Podewski (Hg.), *Text – Material – Medium. Zur Relevanz editorischer Dokumentationen für die literaturwissenschaftliche Interpretation* (Beihefte zu *editio* 37), Berlin/Boston 2014, 25–52.
- Santangelo, Chiara, *Der urheberrechtliche Schutz digitaler Werke. Eine vergleichende Untersuchung der Schutz- und Sanktionsmaßnahmen im deutschen, italienischen und englischen Recht*, Berlin 2011.
- Schmiedt, Helmut, „Tod des Autors und Krise des Werkbegriffs – als Phänomene der U-Kultur“, in: *KulturPoetik* 3/2 (2003), 270–278.
- Spoerhase, Carlos, „Was ist ein Werk? Über philologische Werkfunktionen“, in: *Scientia Poetica. Jahrbuch für Geschichte der Literatur und der Wissenschaften/Yearbook for the History of Literature, Humanities, and Sciences* 11 (2007), 276–344.
- Stobbe, Martin, „Quellcode lesen? Ein Plädoyer für Procedural Literacy in den Literaturwissenschaften“, in: Sebastian Böck/Julian Ingelmann/Kai Matuskiewicz u.a. (Hg.), *Lesen X.0. Rezeptionsprozesse in der digitalen Gegenwart*, Göttingen 2017, 47–68, http://www.v-r.de/_uploads_media/files/9783847107453_boeck_et_al_lesen_wz_082839.pdf.
- Summerer, Claudia, *„Illegale Fans“. Die urheberrechtliche Zulässigkeit von Fan Art*, Berlin/Boston 2015.
- Sutrop, Margit, „The Death of the Literary Work“, in: *Philosophy and Literature* 18/1 (1994), 38–49.
- Theisohn, Philipp, *Literarisches Eigentum. Zur Ethik geistiger Arbeit im digitalen Zeitalter. Essay*, Stuttgart 2012.
- Thomé, Horst, „Werk“, in: *Reallexikon der deutschen Literaturwissenschaft. Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*, Bd. III: P–Z, hg. von Jan-Dirk Müller, gem. mit Georg Braungart u.a. Berlin/New York 2003, 832–834.
- Weiß, Elisabeth, „Wiederkehr des Werks? Zur Gegenwart des literarischen Werkbegriffs“ (Symposium in Hannover v. 21.–23. Oktober 2015), in: *Zeitschrift für Germanistik. Neue Folge* 26/2 (2016), 420–423.
- Winko, Simone, „Am Rande des Literaturbetriebs. Digitale Literatur im Internet“, in: Heinz Ludwig Arnold/Matthias Beilein (Hg.), *Literaturbetrieb in Deutschland*, München 2009, 292–303.
- Wolf, Werner, „Du texte à l’oeuvre? Zur Sinnhaftigkeit der Restauration bzw. Wiederverwendung des Werkbegriffs als eines Grundkonzeptes nicht nur der Literaturwissenschaft“, in: Lutz Danneberg/Annette Gilbert/Carlos Spoerhase (Hg.), *Das Werk. Zum Verschwinden und Fortwirken eines Grundbegriffs*, Berlin/Boston 2019, 379–396.
- Ziegler, Katharina, *Urheberrechtsverletzungen durch Social Sharing. Urheber- und haftungsrechtliche Aspekte sozialer Netzwerke am Beispiel der Plattform Facebook*, Tübingen 2016.
- Zima, Peter V., „Anwesenheit und Abwesenheit des Werks. Zu Foucaults Subjekt- und Werkbegriff“, in: Klaus-Michael Bogdal/Achim Geisenhanslüke (Hg.), *Die Abwesenheit des Werkes. Nach Foucault*, Heidelberg 2006, 181–191.

Online-Ressourcen⁸⁷

- Auer, Johannes/Heibach, Christiane/Suter, Beat (Hg.), *netzliteratur.net. Netzliteratur // Internetliteratur // Netzkunst* (2018), <https://www.netzliteratur.net>.
- Bundesministerium der Justiz und für Verbraucherschutz, *Gesetz über Urheberrecht und verwandte Schutzrechte* (2018), https://www.gesetze-im-internet.de/urhrg/c'est_clairette, <http://cestclairette.com/>.

⁸⁷ Sämtliche Online-Ressourcen wurden letztmalig am 1.5.2018 eingesehen.

DFG-Graduiertenkolleg *Literatur und Literaturvermittlung im Zeitalter der Digitalisierung*, <https://www.uni-goettingen.de/de/kollegiat/innen+und+forschungsstudent/innen/447610.html>.

FanFiktion.de – Das Fanfiktion-Archiv, <https://www.fanfiction.de/>.

Genius, <https://genius.com/>.

Google Books Ngram Viewer, Graph these comma-separated phrases: Daten, Netzwerk, Text, Werk, https://books.google.com/ngrams/graph?content=Werk%2CText%2CDaten%2CNetzwerk&year_start=1720&year_end=2008&corpus=20&smoothing=5&share=&direct_url=t1%3B%2C%20Werk%3B%2C%20c0%3B.t1%3B%2C%20Text%3B%2C%20c0%3B.t1%3B%2C%20Daten%3B%2C%20c0%3B.t1%3B%2C%20Netzwerk%3B%2C%20c0 (letzter Aufruf 22.6.2019).

Herbst, Alban N., *Die Dschungel. Anderswelt*, <http://albannikolaiherbst.twoday.net/>.

Journelles, <https://www.journelles.de/>.

Netzliteraturwissenschaft – Forschung und Community (2021) <https://netzliteraturwissenschaft.net/> (letzter Aufruf 10.09.2021).

Tausend Tode Schreiben, <https://orbanism.com/1000tode/>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Ein Quantum Literatur. Empirische Daten zu einer Theorie des literarischen Textumfangs

Frank Fischer und Robert Jäschke

1 Zum Begriff ‚Textumfang‘

1.1 Einleitung

Ob bewusst oder unbewusst, bei jeder Beschäftigung mit literarischen Texten ist er stetig präsent: der Textumfang. Der vorliegende Aufsatz betrifft damit einen so allgegenwärtigen Gegenstand, dass es verwundern muss, darüber noch keine einzige dedizierte Studie vorzufinden. Immerhin erscheint das Thema allmählich auf der Agenda: So gab es vor Kurzem einen entsprechenden Aufruf zur Bewerbung für ein geplantes Panel mit dem Titel *On Length*, das auf der 2017er Jahreskonferenz der *ACLA (American Comparative Literature Association)* stattfinden und erörtern sollte, „how length – short or long or in between – matters to literary studies. What is the importance of length to literary works?“¹

Nun, das geplante Panel wurde leider für dieses Mal abgelehnt. Dabei wäre eine solche Veranstaltung höchst relevant, denn unsere Kenntnisse über die Semantik literarischer Textumfänge sind allenfalls marginal und kommen bisher höchstens noch als nicht allzu wichtig zu nehmende Größe bei der Unterscheidung

¹Lindsay Thomas/Shannon Brennan, *On Length* (Ausschreibung für ein geplantes, dann aber abgelehntes Seminar für die ACLA-Konferenz 2017), 2016, nur noch über das *Internet Archive* zugänglich: <https://web.archive.org/web/20160921040650/http://acla.org/length> (letzter Aufruf 6.5.2018).

F. Fischer (✉)
Freie Universität Berlin, Berlin, Deutschland
E-Mail: fr.fischer@fu-berlin.de

R. Jäschke
Humboldt-Universität zu Berlin, Berlin, Deutschland
E-Mail: robert.jaeschke@hu-berlin.de

verschiedener Genres ins Spiel, etwa in Aussagen wie dieser: „Nach ihrer Länge steht die Novelle zwischen Roman und Kurzgeschichte.“² Denn, genau, eine Kurzgeschichte ist eben: ‚kurz‘, Novellen meist irgendwie länger, aber selten so lang wie ein Roman, und als ‚Roman‘, als *Novel*, bezeichnet man dann „[a]ny fictitious prose work over 50,000 words“³, um E. M. Forster zu zitieren. Nun hat Marcel Reich-Ranicki dieses launige Bonmot einmal auf seine Weise übersetzt, indem er Forster zu zitieren vorgibt, aber den Roman als „ein erzählendes Werk mit mehr als 200 Seiten“⁴ charakterisiert. Damit sind wir mittendrin: Wir haben es plötzlich mit Zahlen in verschiedenen Einheiten zu tun, von denen eine unveränderlich ist (die Anzahl von Wörtern, aus denen ein veröffentlichter Text besteht), die andere jedoch nicht (je nach Ausgabe eines Werks können sich Seitenanzahlen natürlich voneinander unterscheiden). Diese variabel-invariable Doppelnatur des Textumfangs muss eine entsprechende Theorie fassbar machen.

Sowohl bei der Anzahl der Wörter als auch bei der Anzahl der Seiten handelt es sich im Prinzip um Metadaten, die aber nicht als neutral bibliographisch missverstanden werden dürfen. Der Umfang, egal wo dieser zwischen „short or long or in between“ angesiedelt ist, steht nicht außerhalb der Bedeutung eines literarischen Werks, sondern ist grundsätzlich mit interpretierbar. Wenn wir von literarischen Texten in Buchform reden, materialisieren sich die genannten Metadaten, der Textumfang wird an der Stärke des Buchrückens einschätzbar und ist als Gewicht des Buchblocks auch spürbar, was den Lektüreprozess bereits beeinflussen kann. Dass in der menschlichen Wahrnehmung Gewicht mit Bedeutung korreliert wird, legen entsprechende psychologische Untersuchungen nahe, die zeigen, „that the conceptualization of importance is grounded in bodily experiences of weight“: „People may thus associate the experience of weight with the increased expenditure of bodily or mental effort.“⁵

Wenn man die Zeile als erste (lineare) und die Seite als zweite (flächige) Dimension eines Buches beschreibt, dann liegt es nahe, das Buch als Ganzes als dreidimensionales Objekt aufzufassen, wie dies jüngst Carlos Spoerhase in seiner Studie *Linie, Fläche, Raum* getan hat. Allerdings konstatiert Spoerhase ein „theoretische[s] Desinteresse an der Dreidimensionalität der buchförmigen Schriftlichkeit“,⁶ das er schon bei der einflussreichen buchtheoretischen und bibliophilen Essayistik der 1920er Jahre (mit Valéry, Benjamin und Moholy-Nagy) ausmacht und das bis in die Gegenwart anhält. Spoerhase geht es vor allem um die materialen Aspekte des Textumfangs, also die Implikationen, die

²Gennady Vasilyev, *Wiener Moderne. Diskurse und Rezeption in Russland*, Berlin 2015, 305.

³E. M. Forster, *Aspects of the Novel*, hg. von Oliver Stallybrass, Harmondsworth 1976 [1927], 25.

⁴Marcel Reich-Ranicki, „Wer will, soll’s besser machen“, in: *Die Welt* (21. Oktober 2013).

⁵Nils B. Jostmann/Daniël Lakens/Thomas W. Schubert, „Weight as an Embodiment of Importance“, in: *Psychological Science* 20/9 (September 2009), 1169–1174, hier: 1169.

⁶Carlos Spoerhase, *Linie, Fläche, Raum. Die drei Dimensionen des Buches in der Diskussion der Gegenwart und der Moderne* (Valéry, Benjamin, Moholy-Nagy), Göttingen 2016, 47.

das Buch als sich je spezifisch materialisiert habender Text mit sich bringt. Diese Aspekte spielen in die hier vorliegende Studie hinein, unsere Empirie bezieht sich allerdings auf digital erfasste Metadaten, Katalogdaten letztlich.

Das Interesse an bezifferbaren Textumfangsphänomenen zeigt sich innerhalb der Literaturwissenschaft bisher nur bezüglich kleinerer Einheiten, etwa der Länge von Buchtiteln oder Sätzen.⁷ Aber auch vor der ausdrücklichen Beschäftigung mit dem Umfang literarischer Volltexte ist die Gesamtlänge eines untersuchten Textes eine wichtige *implizite* Größe in der quantifizierenden Literaturwissenschaft, man denke an simple Sachen wie die Berechnung des Vokabelreichtums eines Textes über die Bestimmung der Type-Token-Ratio. Allerdings ist unser Blick auf Textlängen nun ein *expliziter*.

Doch bevor wir mit der Erhebung und Auswertung empirischer Daten beginnen, die als Beitrag zu einer Theorie des literarischen Textumfangs gedacht sind, sollen einige Begrifflichkeiten geklärt und einige Aspekte des bisher nur verstreut geführten Diskurses zusammengeführt werden.

1.2 Textumfang als Teil des Paratextes

Wenn man die materialen Aspekte des Textumfangs einmal außen vor lässt, könnte man die ausdrückliche Angabe von Seitenanzahlen auch als Teil des Paratextes verstehen, also zum „Beiwerk des Buches“ zählen im Sinne Gérard Genettes,⁸ entweder zum Peritext, wenn im Buch selbst, also verlagsseitig, eine Angabe wie z. B. „580 Seiten“ zu finden ist (was so explizit eher selten der Fall ist, denn ein rascher Blick auf die Randbereiche der letzten Buchseiten erfüllt bereits denselben Zweck), oder zum Epitext, wenn diese Angabe über Verlagsprospekte, Rezensionen usw. erfolgt.

Es ist in diesem Zusammenhang übrigens wichtig, zwischen ‚Seitenanzahl‘ und ‚Seitenzahl‘ zu differenzieren.⁹ Erstere ist eine fixe Angabe pro Buch, letztere findet sich auf der Mehrzahl der Seiten eines Buches und etabliert ein numerisches Orientierungs- und Referenzraster. Seitenzahlen wären, obwohl hochkonventionalisiert, noch einmal ein eigenes Thema, denn sie können sich doch auch in die Semantik eines Textes einordnen, wie eine berüchtigte Stelle in

⁷Vgl. Franco Moretti, „Style, Inc. Reflections on Seven Thousand Titles (British Novels, 1740–1850)“, in: *Critical Inquiry* 36/1 (Herbst 2009), 134–158; Stanisław Drożdż/Paweł Oświęcimka/Andrzej Kulig u. a., „Quantifying origin and character of long-range correlations in narrative texts“, in: *Information Sciences* 331 (20. Februar 2016), 32–44; Christof Schöch, „Wiederholende Forschung in den digitalen Geisteswissenschaften“, in: *Konferenzabstracts zur DHD2017 in Bern* (13.–18.2.2017), 207–212.

⁸Gérard Genette, *Seuils*, Paris 1987.

⁹‚Seitenzahl‘ ist, siehe *Duden*, doppeldeutig: „1. Gesamtheit der Seiten eines Druck-Erzeugnisses/2. Zahl, mit der eine Seite eines Druck-Erzeugnisses nummeriert ist“ (<http://www.duden.de/rechtschreibung/Seitenzahl> [letzter Aufruf 6.5.2018]). Aus diesem Grund haben wir uns entschieden, von Seitenanzahlen zu sprechen, wenn wir die Gesamtseitenzahl eines Buches meinen.

Brentanos Roman *Godwi* (1801) zeigt: „Dies ist der Teich, in den ich Seite 266 im ersten Bande falle.“¹⁰

Genette hat die *Seitenzahlen* wohl aufgrund ihrer hohen Konventionalisierung übersehen und die *Seitenanzahl* nicht als Paratext berücksichtigt, weil sie für ihn schlicht zu den rein bibliographischen Metadaten gehört. Und in der Tat oszilliert die Seitenanzahl eines Werkes auch zwischen ihren Funktionen als Teil des Paratextes und Teil der Metadaten. Diese Erkenntnis ist deshalb wichtig, weil sich daraus ergibt, dass der Buchumfang keine bloß bibliothekarische Metaangabe ist, die nicht weiter bedeutungstragend wäre, sondern dass im Kontext des Paratextes dessen Einfluss auf den Rezeptionsprozess diskutierbar wird. Er stellt eine Kommunikationsinstanz dar, der eine Mitteilung, ein Adressant und ein Adressat zugeordnet werden können.

Wie oben bereits gezeigt, handelt es sich beim ‚Textumfang‘ um ein Phänomen, das sich in verschiedenen Formen zeigen kann, sei es als Anzahl von Zeichen, von Wörtern, von Seiten oder als Gewicht des Buchblocks. All diese Größen hängen voneinander ab, in ihrem Verhältnis zueinander gibt es jedoch Spielraum. Um ein Beispiel zu nennen: Die Anzahl von Wörtern in Peter Handkes *Mein Jahr in der Niemandsbucht*¹¹ wird sich zwischen der Erstausgabe von 1994 und der Taschenbuchausgabe von 2007 nicht oder nicht großartig geändert haben, allerdings hat erstere Ausgabe 1.067 Seiten, letztere nur 628.¹² Der wortbasierte Textumfang bleibt unbeeinflusst, der seitenbasierte jedoch nicht. Die Bedeutung dieser doch beträchtlichen materialen Veränderung soll an dieser Stelle noch nicht weiter diskutiert werden, es ging zunächst nur noch einmal um den Hinweis auf die unterschiedlichen Einheiten, in denen sich Textumfang konstituiert. Solcherlei Befunde gehören zur Ausgangslage einer Textumfangstheorie, die sich in diesen rein quantitativen Messungen aber natürlich nicht erschöpfen wird.

2 Diskurssplitter

2.1 Grenzen der Lektüre: Textumfang als Zeitproblem

Der Umfang literarischer Werke ist ein Teil der Form, der die Wirkung eines Textes maßgeblich mitbestimmen kann und verschiedene Rezeptionsaspekte beeinflusst. Zuallererst setzt er der kumulierenden Lektüre enge Grenzen, wie das

¹⁰ Joseph Wälzholz, „Clemens Brentano und das Geheimnis der Seitenzahlen“, in: *Die Welt* (13. September 2014).

¹¹ Peter Handke, *Mein Jahr in der Niemandsbucht. Ein Märchen aus den neuen Zeiten*, Frankfurt/M. 1994.

¹² Raimund Fellinger, „„Schreiben: Sich zur Ruhe setzen“. Die Entstehung von *Mein Jahr in der Niemandsbucht*“, in: Kastberger, Klaus (Hg.), *Peter Handke. Freiheit des Schreibens – Ordnung der Schrift*, Wien 2009, 133–142, 164–173, hier: 141.

einfache Rechenbeispiel zeigt, das Arno Schmidt Mitte der 1950er-Jahre durchgeführt hat:

Das Leben ist so kurz ! Selbst wenn Sie ein Bücherfresser sind, und nur fünf Tage brauchen, um ein Buch zweimal zu lesen, schaffen Sie im Jahre nur 70. Und für die fünf- undvierzig Jahre, von Fünfzehn bis Sechzig, die man aufnahmefähig ist, ergibt das 3150 Bände : die wollen sorgfältigst ausgewählt sein !¹³

Auf ein ähnliches Rechenergebnis kommt vier Jahrzehnte nach Schmidt auch Heinz Schlaffer: „Selbst ein habitueller Leser vermag bestenfalls ein Buch pro Woche zu lesen, im Jahr demnach fünfzig, im Laufe des Lebens vielleicht dreitausend Bücher.“¹⁴ Wiederum knapp anderthalb Jahrzehnte später erhöht Fotis Jannidis gegenüber der *Süddeutschen Zeitung* diesen Wert leicht: „Wir können etwa 4000 Bücher in unserem Leben lesen – das bedeutet, dass jeder von uns eigentlich zu wenig Bücher kennt“.¹⁵ Jannidis zieht diese anthropologisch bedingte Grenze als Grund für die notwendige Digitalisierung der Literaturwissenschaften heran, insofern sich diese jenseits des bekannten Kanons noch am Big Picture versuchen will. Und – so formulierte es ja bereits Moretti – „if you want to look beyond the canon [...], close reading will not do it“.¹⁶ Dieser Gedankengang motivierte dann das *Distant Reading*, für das sich mittlerweile auch eine lebendige digitale Praxis entwickelt hat.

Schmidt und Schlaffer geht es hingegen um Leseökonomie, um Lösungsansätze angesichts des von Schmidt skandalisierten Zeitproblems, um den eng begrenzten „Zeithaushalt des Menschen [...], die verwundbarste Stelle seiner Existenz“, um es mit Hans Blumenberg noch etwas dramatischer auszudrücken.¹⁷ Eine der Konsequenzen, die größere Textumfänge so mit sich bringen, hat dabei in jüngerer Zeit einige Beachtung erfahren, nämlich der Umstand, dass der Normalfall beim Lesen eines Werks nicht dessen Gesamt-, sondern dessen Teillektüre ist, die sich wiederum hochgradig abstufen lässt. Schlaffer unterscheidet im bereits zitierten Text allein 17 Arten der Lektüre, von denen nur die ‚vollständige Lektüre‘ (Typ 12) und die ‚wiederholte Lektüre‘ (Typ 17) implizieren, dass ein Werk ganz gelesen wurde.¹⁸ Auch Pierre Bayard hat sich in seinem paradoxerweise viel gelesenen Bestseller *Comment parler des livres que l'on n'a pas lus*?¹⁹ (2007)

¹³Arno Schmidt, „Ich bin erst sechzig“, in: *Bargfelder Ausgabe*, Werkgruppe I, Band 4. Zürich 1987 [1955], 30 f. (Leer- vor Satzzeichen wie im Original).

¹⁴Heinz Schlaffer, „Der Umgang mit Literatur. Diesseits und jenseits der Lektüre“, in: *Poetica* 31 (1999), 1–25, hier: 5.

¹⁵Boris Hänsler, „Wissenschaft ohne Geist“, in: *Süddeutsche Zeitung* (15. Mai 2013), 16.

¹⁶Franco Moretti, „Conjectures on World Literature“, in: Ders., *Distant Reading*, London/New York 2013 [2000], 43–62, hier: 48.

¹⁷Hans Blumenberg, *Beschreibung des Menschen*, aus dem Nachlaß, hg. von Manfred Sommer, Frankfurt/M. 2006, 621.

¹⁸Schlaffer (Anm. 14).

¹⁹Pierre Bayard, *Comment parler des livres que l'on n'a pas lus*?, Paris 2007.

diesem Aspekt gewidmet, er soll an dieser Stelle jedoch keine weitere Rolle spielen.

Rein rechnerisch ist an den oben mit Schmidts Zitat beginnenden Lesevermögensberechnungen übrigens auffallend, dass in ihnen stets ein wichtiger Faktor fehlt. Die vorgestellten Gleichungen sind nämlich längeninvariant, der Umfang der Bücher spielt keine Rolle. Man benötigt offenbar immer dieselbe Zeit, um ein Buch zu lesen, egal wie lang, egal wie schwierig es ist; Schmidt und Schläffer gehen hier wohl von einem subjektiv empfundenen Mittelwert aus. Textumfang und Lektüredauer stehen natürlich keineswegs immer im gleichen Verhältnis, zumindest die Textschwierigkeit käme als intersubjektiv schwer zu bestimmender Koeffizient noch hinzu, und trotzdem ist es zunächst der Umfang, der die Lektüre steuert. Wir würden behaupten, dass kaum ein Lesevorgang beginnt ohne vorausgehende Orientierungsgeste zur Feststellung, wie lang der bevorstehende Absatz, ein zu lesendes Kapitel oder ein Gesamttext ungefähr ist (den Verzicht auf diese Geste könnte man die Kolumbus-Methode nennen: einfach mal loslesen und schauen, wohin man kommt). Als Beispiel für neuere Entwicklungen in Sachen Textorientierung seien die variable Größe der Browserleiste und die Prozentangaben bei E-Books genannt, und „[i]m Onlinejournalismus ist es inzwischen gang und gäbe, Beiträgen einen Hinweis auf die voraussichtliche Lektüredauer voranzustellen“.²⁰ Wobei dies, im Grunde genommen, gar nicht so neu ist, ein prominenter Vorläufer wäre etwa Klabunds Hundert-Seiten-Buch *Deutsche Literaturgeschichte in einer Stunde* (1920), dessen Titel trotz der ironischen Brechung durchaus ganz konkret gemeint ist.

2.2 Mittlere und vorgegebene Textumfänge

Eine Abkürzung anderer Art schwebte Hans Magnus Enzensberger vor. In dem *Ideen-Magazin*, das seine Anfang 2011 erschienene autobiographische Sammlung *Meine Lieblings-Flops* beschließt, stellt er Projekte vor, die „über das Stadium der Skizze nie hinausgekommen“ sind. Eines davon trug den Arbeitstitel *Die hundert Seiten*: Klassiker der Weltliteratur, die besonders umfangreich, besonders unzugänglich sind, sollten auf genau hundert Seiten komprimiert werden, als Nacherzählung eigenen Rechts. Denn „viele der berühmtesten Klassiker werden nicht gern gelesen“, da sie eine „Zumutung an das Zeitbudget“ seien.²¹ Er siedelte

²⁰Mark-Stefan Tietze, „Lesedauer: 100 min“, in: *die tageszeitung* (13. Juni 2017). Dass diese neueren Orientierungsmethoden als Alternative zur Buchseite nicht automatisch eine selbsterklärende kulturelle Praxis sein müssen, illustriert diese Kundenfrage aus dem *Kindle Hilfeforum* von Amazon: „Wenn ich dann das Buch aufschlage steht unten immer ‚1 h 34 min verbleiben für das Buch‘. Soll das etwa heißen, ich muss das Buch innerhalb dieser Zeit durchgelesen haben?“ [sic] Anja H., 22. Dezember 2014, <https://www.amazon.de/gp/help/customer/forums/kindleqna/?cdThread=Tx36D9B96YMGG8S> (letzter Aufruf 6.5.2018).

²¹Hans Magnus Enzensberger, *Meine Lieblings-Flops, gefolgt von einem Ideen-Magazin*, Berlin 2011, 233.

sein Projekt selbst im literarischen Bereich an, die Zusammenfassungen sollten wiederum von Romanciers verfertigt werden, so entstandene Kurzversionen also literarisch eigenständig sein und sich darin von Zusammenfassungen etwa in Literaturlexika unterscheiden.

Das Lamento über zu lange Bücher kann auch bei den Autoren selbst zur Mäßigung führen. Zwei argentinische Autoren seien angeführt, die daraus gar eine Programmatik entwickelt haben. Jorge Luis Borges etwa meint im Prolog zu dem Band *El jardín de senderos que se bifurcan* (1941), dass es ein mühsamer und auslaugender Unsinn sei, dicke Bücher zu schreiben und auf mehr als 500 Seiten einen Gedanken auszubreiten, der mündlich in wenigen Minuten akkurat mitgeteilt sei.²² Auch bei César Aira findet sich diese Präferenz für kurze Texte. In der Erzählung *El Congreso de Literatura* (1997) lässt Aira seinen Erzähler, der den gleichen Namen trägt wie der Autor, sagen: „me he impuesto una extensión fija para todo el texto [...] por respeto al tiempo del lector“, dass er es sich also auferlegt hat, beim Gesamttext eine gewisse Seitenzahl nicht zu überschreiten, aus Respekt vor der (begrenzten) Zeit des Lesers.²³ Aira hat in mehreren Interviews unterstrichen, dass es ihm beim Verfassen seiner meist um die 100 Seiten langen Erzählungen bzw. Romane nur darauf ankommt, wenigstens Buchlänge zu erreichen.²⁴ Koketterie oder nicht, zeitökonomische Überlegungen können auch Teil poetologischer Konzepte sein.

Eine weitere Möglichkeit der kontrollierten Einhegung des Textumfangs ist dessen von vornherein Fixierung. Für die Einhaltung einer festen Länge gibt es auch in der fiktionalen Literatur Beispiele, wenn auch eher im U-Bereich und im Rahmen entsprechender Reihen, man denke etwa an den 64-seitigen Heftroman.

Doch wann ist ein Textumfang ‚perfekt‘? Das kann man fragen, wenn man „length“ auffasst als „an aspect of any book that has an enormous if somewhat mysterious influence on its quality“.²⁵ Schon in der aristotelischen *Poetik* finden sich Bemerkungen zur perfekten Länge (der „richtige[n] Begrenzung der Ausdehnung“) des Epos.²⁶ Dem zugrunde liegt die Annahme, dass es für die Ausgestaltung eines geeigneten Stoffes jeweils die perfekte Länge gibt – eine Annahme, die sich beim modernen Roman natürlich erübrigt hat. Und doch pegeln sich bestimmte Textumfänge ein und steuern die Erwartungshaltung. Ein Beispiel

²² Jorge Luis Borges, *Ficciones – El Aleph – El Informe de Brodie*, Caracas 1993, 5.

²³ Aira César, *El Congreso de Literatura*. Mérida: Fundación Casa de las Letras „Mariano Picón Salas“, Mérida 1997, 12.

²⁴ Ernesto Escobar Ulloa, „Entrevista a César Aira“, in: *The Barcelona Review. Revista Internacional de Narrativa Breve Contemporánea* 44 (Sept./Okt. 2004).

²⁵ Laura Miller, „Are longer books more important?“, in: *Salon.com* (6. September 2012).

²⁶ Aristoteles, *Poetik*, Griechisch/Deutsch, übers. und hg. von Manfred Fuhrmann, Stuttgart 1997 [1982], 81.

nennt der Verleger Michael Krüger: „Es gibt eine Euro-Novel: 320 Seiten, moderates Thema, das hat sich so eingespielt.“²⁷

Genau dieser Textumfang, die 320 Seiten, spielte vor zwei Jahrhunderten in einem ganz anderen Kontext schon einmal eine Rolle. Diese Anzahl von Seiten (die 20 Bogen entspricht) musste laut den Karlsbader Beschlüssen von 1819 überschritten werden, damit ein Buch nicht der obligatorischen Vorzensur unterlag (der Gedanke dahinter war, dass sich der Inhalt dicker Bücher schlechter verbreiten ließe). Nach dieser 20-Bogen-Klausel war es also möglich, kritische Schriften erst einmal in Umlauf zu bringen, insofern nur ein gewisser Umfang erreicht werden konnte.²⁸ So war etwa Heines *Deutschland. Ein Wintermärchen* (1844) in einem Sammelband erschienen, da es allein zu kurz gewesen wäre, um die 20 Bogen zu erreichen.

2.3 Extremtexte: „Der längste Roman der Welt“

Eine weitere Art, sich dem Thema ‚Textumfänge‘ zu nähern, ist die Fokussierung auf Texte, die sich in umfangstechnischen Extrembereichen bewegen. Dies kann zunächst auch das Niedrigstsegment sein, man denke an berühmte Kürzesttexte wie Apollinaires Einzeiler *Chantre* aus der 1913 erschienenen Sammlung *Alcools* („Et l’unique cordeau des trompettes marines“) und Ungarettis trotz Zweizeiligkeit noch kürzeres Gedicht *Mattina* („M’illumino/d’immenso“, 1917) oder auch Daniil Charms’ Zweisatzerzählung *Vstreča* (vor 1942), die allesamt dadurch frappieren, dass sie die genrebedingte Mindestervartung an den Umfang eklatant unterlaufen.

Die „literarischen Formate des Kleinen“ haben jüngst dezidiert Beachtung gefunden, nicht zuletzt, da sie dem Zeitgeist entgegenzukommen scheinen, denn die Literatur reagiere im 20./21. Jahrhundert auf die „Begrenzung von Aufmerksamkeits- und Zeitressourcen mit dem ‚kommunikativen Imperativ‘ der Kürze“, um die Ankündigung für ein entsprechendes Symposium an der *Universität Paderborn* zu zitieren.²⁹ Seit dem Frühjahr 2017 widmet sich auch ein in Berlin beheimatetes *DFG-Graduiertenkolleg* dem Thema (*Literatur- und Wissenschaftsgeschichte kleiner Formen*).

Bei der Diskussion um Texte extrem großen Umfangs sieht es etwas anders aus. Wenn man von Ausnahmen wie Stefano Ercolinis [2014](#) erschienener Studie

²⁷ Christian Zschke, „Klopfer für die Ewigkeit“, in: *Süddeutsche Zeitung* (16. Dezember 2009), 3.

²⁸ Vgl. Jost Schneider, *Sozialgeschichte des Lesens. Zur historischen Entwicklung und sozialen Differenzierung der literarischen Kommunikation in Deutschland*, Berlin/New York 2004, 168.

²⁹ Leonie Süwolto, „Pragmatik und Ästhetik des Kleinen. Literarische, visuelle und mediale Mikroformate im 20. und 21. Jahrhundert“ [Konferenzankündigung], in: *H/Soz/Kult* (15. März 2013).

*The Maximalist Novel*³⁰ absieht (wobei „length“ bei ihm nur eine von zehn Eigenschaften des maximalistischen Romans ist), haben unseres Wissens bisher keine größeren Anstrengungen stattgefunden, um die Extreme literarischen Umfangs dezidiert zu untersuchen. Normalerweise, vor allem abseits des Wissenschaftsbetriebs, beschränkt sich das Interesse daran auf eine Superlativik, die letztlich nur auf sich selbst verweist. Das beginnt mit dem *Guinness-Buch der Rekorde* und dem Gewinner der Rubrik *Longest novel*.³¹ Die Nase vorn, sozusagen, hat Prousts *À la recherche du temps perdu* – und zwar mit „an estimated 9,609,000 characters (each letter counts as one character. Spaces are also counted, as one character each).“³²

Diese Angabe ist aus mehreren Gründen problematisch. Erstens, weil nicht thematisiert wird, was als *Novel* zählt (die romans-fleuves und anderweitige Romanzyklen? und ab wann ist ein Roman kein Fortsetzungsroman mehr? usw.). Und zweitens, weil es natürlich sehr wohl längere Romane gibt.

Der beliebteste Treff für die Diskussion um die längsten Romane der Welt ist aber der Artikel *List of longest novels* in der englischen Wikipedia.³³ Er hält eine Liste mit Werken einer Länge von „over 500,000 words published through a mainstream publisher“ bereit und wird mit relativ hoher Frequenz editiert. Hier findet sich dann gleich im Artikelkopf auch die vermeintlich richtige Antwort auf die Frage nach dem längsten Roman. Anhand der kursierenden Seitenumfangszahl 13.095 wird dazu Madeleine de Scudéry's Roman *Artamène, ou le Grand Cyrus* erklärt, erschienen in zehn Bänden zwischen 1649 und 1653 (wohlgemerkt, es handelt sich nicht um das längste ‚Buch‘).

Im präsentierten Ranking allerdings findet sich dann ein roman-fleuve an der Spitze, *Les Hommes de bonne volonté* von Jules Romains, erschienen in 27 Bänden zwischen 1932 und 1946. Die Community sieht ihn bezüglich der Zeichenanzahl vor Madeleine de Scudéry, und es ist so sinnlos wie eigentlich unmöglich, hier irgendwelche abschließenden Urteile zu fällen, was auch der Grund dafür ist, dass der *Wikipedia*-Artikel seit seiner Entstehung im Jahr 2004 mehrfach tiefgreifend umstrukturiert wurde. Als längster deutscher Roman wird übrigens *Zettel's Traum* geführt, aber auch dies ist natürlich kontingent, und nennen müsste man hier wenigstens noch die *Römische Octavia* des Herzogs Anton Ulrich von Braunschweig-Wolfenbüttel, wobei das Zählen der Seiten bei diesem über mehrere Jahrzehnte erschienenen und mehrfach umgearbeiteten

³⁰Stefano Ercolino, *The Maximalist Novel. From Thomas Pynchon's Gravity's Rainbow to Roberto Bolaño's 2666*, New York/London 2014.

³¹O.N., „Longest novel“, in: *Guinness World Records*, <http://www.guinnessworldrecords.com/world-records/longest-novel> (letzter Aufruf 6.5.2018).

³²Ebd.

³³Siehe *Wikipedia*, https://en.wikipedia.org/wiki/List_of_longest_novels (letzter Aufruf 6.5.2018).

Barockroman eine besondere Schwierigkeit darstellt, doch sind es sicher mehr als 7.000.³⁴

Die *Wikipedia*-Diskussion um den längsten Roman der Welt wird jedenfalls durchaus ernsthaft geführt, dafür spricht die Vielzahl von Kriterien. So sind etwa „self-published, printed-on-demand, vanity works, [...] and record-grabbing stunts“ von der Aufnahme in die Liste ausgeschlossen. Da aber die ausschlaggebenden Kriterien, wie oben angeführt, nicht zu eindeutigen Ergebnissen führen, bleibt die Liste vorläufig. Ihr größtes Problem ist natürlich, dass sie nur auf sich selbst verweist und sich völlig in ihrem Interesse am Superlativ und am Ranking erschöpft.

Der logische nächste Schritt wäre, einmal nach Funktionen des Umfangs dicker Bücher zu fragen. So ist die schiere Präsenz eines massiven Buchblocks zunächst mitunter respektgebietend, wie diese Bemerkung in den *Blättern für literarische Unterhaltung* von 1863 zeigt:

Dieser Respect vor recht dicken Büchern besteht im Grunde auch noch jetzt; man schreibe das geistreichste Buch über Shakspeare [sic], concentrirte aber seine Ansichten in einem dünnen Bande, und die Schrift wird wahrscheinlich wenig Beachtung finden; man dehne es aber zu drei oder vier starken Bänden aus, und das gelehrte und halbgelehrte Publikum wird bewundernd davor stille stehen wie vor einer der ägyptischen Pyramiden, an denen man im Grunde vorzugsweise doch nur die Masse anstaunt.³⁵

Dickleibige Werke verursachen also Bewunderung und laden offenbar gleichzeitig zur Nichtlektüre ein. Fritz J. Raddatz spricht in so einem Fall dann von einem „Werk-Gerücht – Proust, Joyce, ungelesen, aber zum obligaten Kanon zählend“.³⁶ In beiden Zitaten wird suggeriert, dass es einen Zusammenhang geben könnte zwischen extremem Textumfang und Kanontauglichkeit. Ob man nachweisen kann, dass dickleibige Bücher in literarischen Kanons überrepräsentiert sind, soll exemplarisch in Abschn. 4.3. *Textumfang und Kanonisierung* untersucht werden.

Ein anderes Subphänomen im Zusammenhang mit literarischen Texten extremer Ausdehnung soll hier zumindest angeführt werden, auch wenn dessen Erörterung eher in eine rezeptionsästhetische Studie gehört. Man könnte, analog zum Titel von Peter Brooks' Monographie *Reading for the Plot* (1984), von *reading for the length* sprechen, denn der „Respect vor recht dicken Büchern“ geht nicht zwangsweise mit deren Nicht- oder nur Teillektüre einher. In Uwe Tellkamp

³⁴Vgl. Peter Heßelmann, „Herzog Anton Ulrichs Roman *Römische Octavia* in neuer Perspektive“, Rezension zu: Stephan Kraft: Geschlossenheit und Offenheit der „Römischen Octavia“ von Herzog Anton Ulrich. „Der roman macht die ewigkeit gedencken, den er nimbt kein endt“, Würzburg 2004, in: *IASOnline* (5. November 2004).

³⁵O.N., „Notizen. Schriftstellerklagen aus den sechziger Jahren des vorigen Jahrhunderts“, in: *Blätter für literarische Unterhaltung* 27, Bd. 2: Juli bis December (10.9.1863), 682 f., hier: 682.

³⁶Fritz J. Raddatz, *Tagebücher 2002–2012*, Reinbek b. Hamburg 2014, 655.

Roman *Der Turm* (2008), selbst ein Seitenanzahlenschwergewicht, finden sich die Sätze: „Mit 500 Seiten begannen die wirklichen Romane. Mit 500 Seiten begann der Ozean, drunter war Bachpaddeln.“³⁷ Ähnlich äußert sich der Feuilletonist Andreas Plathaus anlässlich seiner Rezension des ebenfalls sehr umfangreichen Romans *Morbus Fonticuli* (2001) von Frank Schulz:

Eines muß der Rezensent wohl vorausschicken: Er liebt lange Filme und dicke Bücher. Nicht daß kurzes Kino und schlanke Literatur keine Gnade vor seinen Augen fänden, aber alles, was hundertachtzig Minuten oder sechshundert Seiten überschreitet, hat es bei ihm leichter.³⁸

Umfang als Gütesiegel findet sich auch bei Christian Kracht, der mit seinen eigenen „dünnen Büchern“ hadert und einem Feuilletonisten der *Welt* erklärt, dass „dicke Bücher [...] nun mal besser als dünne [sind]. ‚Die Enden der Parabel‘ ist besser als ‚Der Fänger im Roggen‘.“³⁹

Doch wie gesagt, diese Überlegungen gehören in einen anderen Subdiskurs und sollen hier nicht weiterverfolgt werden. Insgesamt sollte jedoch etabliert sein, dass die quantifizierbare Größenordnung eines literarischen Textes ein konstitutives formales Merkmal desselben darstellt und stets auch Einfluss auf die ästhetische Erfahrung ausüben kann, auf die ‚Stimmung‘ (im Sinne Hans Ulrich Gumbrechts). Denn „[a]usnahmslos alle Konstitutionsebenen von Texten können an der Produktion von Stimmungen beteiligt sein“,⁴⁰ schreibt Gumbrecht, und dies gilt eben auch für den implizit empfundenen oder explizit präsenten Textumfang.

2.4 ‚Seitenpolitik‘

Der Diskurs über den Textumfang hat nur Sinn, wenn man ‚Umfang‘ als Vergleichsgröße versteht, eine einzige derartige Angabe zu einem einzigen Werk ist nicht mehr als eine bedeutungslose Metainformation. Und Vergleichsaspekte gibt es einige. Man kann etwa dazu ansetzen, „das Werk eines Schriftstellers rein quantitativ zu fixieren und mit anderen vergleichbar zu machen“, wie Arno Schmidt dies vorhatte (dazu unten mehr). Man kann Textumfang genreintern thematisieren, man kann Originale mit Übersetzungen ins Verhältnis setzen und dergleichen mehr, und man kann Umfang auch als Teil von Verlagsstrategien

³⁷ Uwe Tellkamp, *Der Turm. Geschichte aus einem versunkenen Land*, Frankfurt/M. 2008, 154.

³⁸ Andreas Plathaus, „Kindisch klingt der Ruf des Ku[c]kucks“, in: *Frankfurter Allgemeine Zeitung* (1. Dezember 2001), V.

³⁹ Jan Küveler, „Einmal quer durch Faserland“, in: *Welt am Sonntag* (20. November 2016), 57 f., hier: 58.

⁴⁰ Hans U. Gumbrecht, *Stimmungen lesen. Über eine verdeckte Wirklichkeit der Literatur*, München 2011, 13.

erörtern, hier speziell als etwas, das man ‚Seitenpolitik‘ nennen könnte. Die Präsentationsmöglichkeiten eines Textes in einem Buch sind vielgestaltig, verschiedene Satzspiegel und Methoden der Textstauchung und -streckung kamen und kommen zur Anwendung, aus verschiedensten (ökonomischen, ästhetischen, produktionsbedingten usw.) Gründen.

Kommen wir auf ein oben bereits angeführtes Beispiel zurück: Peter Handkes Buch *Mein Jahr in der Niemandsbucht*, das bei seinem ersten Erscheinen 1994 vom Verlag mit einem sehr luftigen Satzspiegel versehen wurde, der nur 23 Zeilen pro Buchseite vorsieht.⁴¹ Nur deshalb kann das Buch auf über 1.000 Seiten Länge anwachsen und ein Tausendseiter sein, und auch nur deshalb kann es vom Erzähler als „eins dieser unanständig dicken Bücher“ etikettiert werden.⁴² Wenn am Umfang ersichtlich ist, dass dieses Werk sich selbst als ‚Großes Werk‘ versteht, als eine *Maximalist Novel*, muss man auch nicht mehr darauf pochen, dies in der Vorlagsvorschau oder dem Klappentext explizit anzukündigen, wie es Handkes Lektor Raimund Fellingner ursprünglich tun wollte: „In seinem großen, neuen, bisher umfangreichsten [...] Werk erzählt Peter Handke [usw.]“ – Handke war entsetzt und beklagte sich sofort bei seinem Verleger Siegfried Unseld, und zwar wie folgt: „Alles, was ich unbedingt vermeiden wollte – daß gesagt würde: ‚groß‘, ‚großes Werk‘, ‚großes Epos‘ –, steht nun großkotzig da.“⁴³ Und das braucht es tatsächlich nicht, denn die angepeilte ‚Größe‘ dieses Romanwurfs wird ja schon durch seinen Status als Tausendseiter kommuniziert. In der späteren Taschenbuchausgabe, siehe oben, wurde der Text auf 628 Seiten zusammengezurr, was man kaum noch als „unanständig dick“ bezeichnen kann.

Neben solchen Einzelfällen kann man sich auch ansehen, welche Extremtexte die Verlage in ihren Programmen platziert haben. Denn wie schon aus dem *Wikipedia*-Artikel zu den längsten Romanen deutlich wird, gibt es gar nicht so viele überlange Romane, die die nötigen 500.000 Wörter aufbringen würden.⁴⁴ Da ‚Buchumfang‘ auch eine ökonomische Komponente hat und lange Bücher das verlegerische Risiko steigern, kann es aussagekräftig sein, die umfangreichsten Texte verschiedener Verlage gegeneinander zu halten, was wir in 4.2. *Die umfangreichsten Bücher pro Verlag* demonstrieren werden.

Wir konnten hier nur einige mögliche Aspekte des beginnenden Diskurses um den Textumfang als literaturwissenschaftliche Kategorie benennen und

⁴¹ Vgl. Fellingner (Anm. 12), 141.

⁴² Handke (Anm. 11), 107.

⁴³ Fellingner (Anm. 12), 141.

⁴⁴ Eine Liste, die sich nicht an der Anzahl von Wörtern, sondern an der Anzahl von Seiten orientiert und dezidiert ‚Tausendseiter‘ sammelt, findet sich in Paco [d. i. Frank Fischer], „Tausendseiter“, in: *Der Umblätterer* (21. Oktober 2014), <http://www.umblaetterer.de/2014/10/21/tausendseiter/> (letzter Aufruf 6.5.2018).

diskutieren, und bei Weitem nicht alle verlangen nach einem empirisch-quantifizierenden Instrumentarium. Da wir nun aber bereits einige dahingehende Versprechungen gemacht, jedoch noch nicht über deren Operationalisierung gesprochen haben, widmen wir uns nun dieser.

3 Operationalisierung

3.1 Katalogbasierte Seitenanzahlen als Maßeinheit des Textumfangs

Angesichts der durchaus variablen Größe ‚Seitenanzahl‘ (siehe das Beispiel Handke), aber auch angesichts der schriftgrößenvariablen Textpräsentation, die sich beim Lesen von E-Books, im Browser usw. findet, wäre ein Messwert jenseits der Angabe von Seitenanzahlen nötig. Was wäre so ein Maß, das Vergleichbarkeit herstellt? In einer für die Publikation gestrichenen Passage aus *Fouqué und einige seiner Zeitgenossen* (1958/1960) stellt Arno Schmidt folgende Überlegungen an:

Wie irreführend ist es oft, zu sagen, ein Buch zähle 500 Seiten; nachher hat es auf jeder einzelnen davon nur 20 Zeilen und in jeder 40 Anschläge = 800 Buchstaben. Ein anderes, von „nur“ 200 Seiten, aber mit 40 Zeilen [à] 50 Anschläge, enthält genau so viel Text. Man führe endlich in Wissenschaft und Buchhandel den Begriff der „Normalseite“ (abgekürzt: SN) von 2000 Buchstaben pro Seite ein! Es bleibe natürlich auch in Zukunft jedem unbenommen, mit Format, Zeilenzahl oder Typen völlig souverän zu schalten, aber man füge der Anzeige auch des apartesten Sonderdruckes noch in Klammern hinzu: „SN 340“ – oder wieviel es nun gerade sind. Das würde, konsequent durchgeführt nicht nur in Katalogen aller Art, viel nützen, sondern endlich auch einmal ermöglichen, das Werk eines Schriftstellers rein quantitativ zu fixieren und mit anderen vergleichbar zu machen.⁴⁵

Zwar hat sich inzwischen in verschiedenen Kontexten der Begriff der ‚Normseite‘ etabliert, doch ist die dahinterstehende Zeichenzahl als Richtwert national und international schwankend. Nichtsdestotrotz ist Schmidts Idee, eine normierte Umfangsangabe für literarische Werke zu schaffen, tatsächlich geeignet, um ‚Umfang‘ als trägermedienundabhängige komparatistische Größe überhaupt nutzbar zu machen. Inwiefern Absätze, Zeilenumbrüche usw. in die Berechnung dieser ‚Normalseite‘ hineinspielen sollen, wäre noch festzulegen, und was eine ‚Normalseite‘ im Fall von Schmidts eigenem Romanungetüm *Zettel's Traum* bedeutete, das ja in drei parallel nebeneinander platzierten Erzählsträngen gesetzt ist, müsste man dabei gleich mit klären.

Wenn wir also im Folgenden mit katalogbasierten Seitenanzahlen rechnen, nehmen wir die etwas problematische Buchseite als Maß der Dinge – eine

⁴⁵ Arno Schmidt, „*Fouqué*“-Typoskript, DLA Marbach, zit. n. Stephan Reuter, *Arno Schmidt und die Bücher – betrachtet aus der Perspektive einer gestrichenen Textstelle der Fouqué-Biographie*, Webpublikation [2000].

Einschränkung, die buchhistorisch allerdings durchaus Sinn hat, wie wir gesehen haben, denn der Diskurs über die Länge literarischer Texte ist noch immer vor allem seitenorientiert.

3.2 „362.597 Seiten“ – Vorarbeit auf Kanon-Ebene

Zum eigentlichen Untersuchungsgegenstand für diese Studie haben wir den Katalog der *Deutschen Nationalbibliothek* erkoren. Da dieser mit seinen mehreren Millionen Medien und einer durchaus teils problematischen Metadatenqualität ein sehr komplexes Thema darstellt, sei hier eine kurze Vorarbeit an einem kleineren Korpus eingeschoben, um überhaupt einmal zu evaluieren, inwiefern Textumfangsdaten ausgewertet und visualisiert werden können.

Geschehen soll dies am Beispiel eines jüngeren (welt)literarischen Kanons, den *1001 Books You Must Read Before You Die*, erschienen 2006, herausgegeben von Peter Boxall.⁴⁶ Die Version von 2008 wurde stark aktualisiert, die weiteren Ausgaben von 2010 und 2012 beinhalten nur marginale Revisionen; wir greifen hier auf die Erstausgabe von 2006 zurück. Es handelt sich wohlgerne um einen „populären Leseführer“,⁴⁷ nicht um einen literaturwissenschaftlichen Kanon.

Unser Freund William McComish hat im Jahr 2012 jedem der 1.001 Bücher seine Seitenanzahl zugewiesen.⁴⁸ Obwohl er dabei nicht die jeweils benutzte Ausgabe spezifiziert hat, geben die Daten einen guten ersten Einblick in die statistische Verteilung der Buchumfänge (die zugrunde liegende TSV-Datei kann in unserem Repositorium eingesehen werden). Insgesamt haben wir es mit 362.597 Seiten zu tun, deren Verteilung auf die tausendundeins Bücher Abb. 1 zeigt.

Auf der *x*-Achse finden sich die Seitenanzahlen als Umfangskategorien in 16er-Schritten (warum es 16er-Schritte sind, dazu unten mehr), und jedes Buch wird nun in seine jeweilige Kategorie einsortiert, sodass die Säulen entlang der *y*-Achse entsprechend wachsen. Das Gros der Bücher ist zwischen 100 und 400 Seiten lang (682 Stück, somit mehr als zwei Drittel). 23 Werke sind länger als 1.000 Seiten, sie wurden in Abb. 1 gemeinsam aufsummiert, obwohl sie sich eigentlich viel weiter in die *x*-Achse hinein verteilen, bis zu einer Seitenanzahl von etwa 3.700, erzielt von dem südkoreanischen roman-fleuve *The Taebaek Mountains* des Autors Jo-Jung Rae. Das kürzeste Buch ist wiederum Gogols Erzählung *Die Nase*, das als alleiniger Gegenstand mit den 32 hier veranschlagten Seiten nicht einmal den

⁴⁶Boxall, Peter (Hg.), *1001 Books You Must Read Before You Die*, London 2006.

⁴⁷Hermann Dorowin, „Zu kurz für den Kanon? Zum Problem der ‚kleinen Form‘“, in: Sanna, Simonetta (Hg.), *Der Kanon in der deutschen Sprach- und Literaturwissenschaft*, Bern u. a. 2009, 137–144, hier: 137.

⁴⁸Vgl. Baumanski [d. i. Fabian Baumann], „362.597 Seiten“, in: *Der Umblätterer* (20. August 2012), <http://www.umblaetterer.de/2012/08/20/362597-seiten/> (letzter Aufruf 6.5.2018).

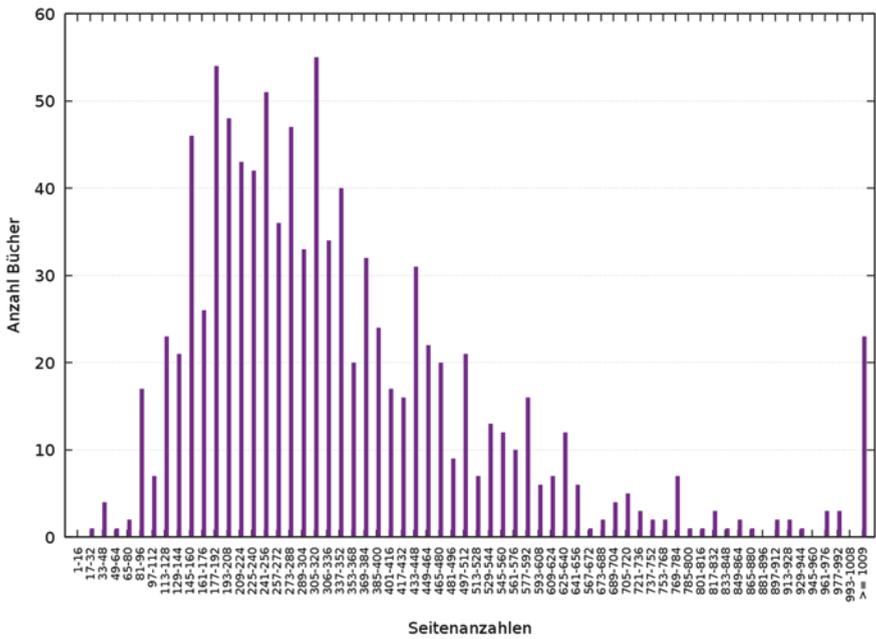


Abb. 1 Verteilung der Seitenanzahlen innerhalb des 1001-Books-Kanons

Tatbestand ‚Buch‘ erfüllt, denn laut UNESCO ist ein Buch „a non-periodical printed publication of at least 49 pages“⁴⁹.

Was kann man nun aus dieser visualisierten Kanonliste ersehen? Zunächst kommt mit der längenbasierten Sortierung des vorgestellten Kanons ein neues Ordnungskriterium ins Spiel, Textumfang wird einmal jenseits der Genrediskussion fassbar und erzeugt etwa einen Subkanon hundertseitiger Bücher – eine Textlänge, für die sich nicht nur, wie oben beschrieben, Enzensberger und César Aira begeistern. Für dieses Subphänomen, das 100-Seiten-Buch, haben wir diese neue Ordnungsmöglichkeit sogar einmal durchgespielt.⁵⁰ Auch hier zeigte sich die Schwankungsbreite, die mit der ‚Seitenpolitik‘ der Verlage einhergeht, denn 100-Seiten-Bücher oszillieren unserer Beobachtung nach in ihrem Umfang zwischen etwa 75.000 und 225.000 Zeichen.⁵¹ Das 100-Seiten-Genre hat übrigens sogar zu entsprechenden verlegerischen Unternehmungen geführt: in Italien zu der

⁴⁹UNESCO, *Recommendation concerning the International Standardization of Statistics Relating to Book Production and Periodicals*, Paris 1964.

⁵⁰Vgl. Frank Fischer u. a., „Lob des 100-Seiten-Buchs“, in: *EDIT. Papier für neue Texte* 61 (Frühjahr 2013), 117–129.

⁵¹Ein *Kanon hundertseitiger Bücher*, der auf wiederum hundert Seiten hundert Vertreter dieses umfangsbasierten Genres präsentiert, ist in Vorbereitung. Vgl. die Liste auf <http://www.umblaetterer.de/100-seiten/> (letzter Aufruf 6.5.2018).

von Italo Calvino herausgegebenen Reihe *Centopagine*, in Frankreich zur *Éditions Cent Pages*. Und auch der Reclam-Verlag hat jüngst eine eigene Reihe namens *100 Seiten* eingeführt, kurze Monographien zu aktuellen Themen, Slogan: „100 Seiten für 100 min“.⁵²

Doch Textumfang als Ordnungsprinzip (das ja z. B. auch die Anordnung der Suren im *Koran* organisiert) hat, für sich besehen, noch keinen unmittelbaren Erkenntniswert. Interessanter wäre da die Frage, ob sich etwas über den Einfluss von Buchlängen auf die (Nicht-)Kanonisierbarkeit von Werken ableiten lässt. Dazu kommen wir auch gleich, wir benötigen nur zunächst eine Vergleichsgröße.

3.3 Der DNB-Katalog

Es gibt noch kaum literaturwissenschaftlich relevante Studien, die den Katalog der *Deutschen Nationalbibliothek (DNB)* als Grundlage für statistische Untersuchungen nutzen.⁵³ Mittlerweile ist die Arbeit mit diesem Katalog sehr komfortabel geworden, der Datendienst der *DNB* stellt vierteljährlich einen Komplettabzug der Katalogdaten im RDF-Format bereit, und zwar unter der freien Lizenz CC0 1.0.⁵⁴ Der von uns verwendete Abzug stammt vom 23. Juni 2017 und enthält 14.102.309 Datensätze. Nicht alle Datensätze repräsentieren Bücher, und nicht alle Bücher weisen extrahierbare Informationen zu ihrem Umfang auf (die meisten aber schon). Man muss außerdem hinzufügen, dass nicht bei allen Büchern mit Umfangangaben diese auch korrekt sind.⁵⁵ Alle für uns relevanten Datenfelder (etwa Autor, Titel, Verlag, Publikationsjahr, Umfang) wurden extrahiert und ggf. normalisiert. Mit dem von uns gebauten Framework ist es so auch möglich, in Detailbereiche des Problems vorzudringen und auf Basis von Textumfangsdaten etwa die Verlagspolitik hinsichtlich des Publizierens seitenanzahltechnisch extremer Werke zu beschreiben.

⁵²Vgl. <https://www.reclam.de/100Seiten> (letzter Aufruf 6.5.2018).

⁵³Eine Ausnahme bilden etwa Günter Häntzschel/Adrian Hummel/Jörg Zedler, *Deutschsprachige Buchkultur der 1950er Jahre. Fiktionale Literatur in Quellen, Analysen und Interpretationen*, Wiesbaden 2009.

⁵⁴Datendienst *Bibliografische Dienstleistungen*, *DNBTitel.rdf.gz* (14504779 Datensätze/235578087 Tripel/DNB-Titel-Daten im Format RDF (RDF/XML), Stand: Juni 2017 (erzeugt am 23. Juni 2017), <http://www.dnb.de/datendienst>).

⁵⁵Das „umfangreichste“ Werk war bis Mai 2016 sage und schreibe 2.711.111 Seiten lang, natürlich ein Verschreiber, der nach unserem Hinweis auch umgehend von der *DNB* korrigiert wurde (vgl. Frank Fischer/Ubbo Veenster, „Empirical Data on Over-Length Books“, in: *weltliteratur.net* (25. Mai 2016); vgl. auch die Antwort der *DNB* auf *Twitter*: https://twitter.com/DNB_Aktuelles/status/736086931646205952 [letzter Aufruf 6.5.2018]). Ein Werk mit mehr als einer Million Seiten hätte natürlich einen störenden Einfluss auf die Auswertung, allerdings bewegen sich noch bestehende Ausreißer deutlich unterhalb dieser Größenordnung (eine Übersicht über die fälschlicherweise „längsten“ Werke findet sich in unserem Repository). Da es sich bei Büchern mit angeblich mehr als 5000 Seiten durchweg um fehlerhafte Angaben handelt, haben wir diese aus unseren Berechnungen herausgenommen, was den Einfluss von Fehlern signifikant eindämmt.

Tab. 1 Die zehn häufigsten Werte des Datenfeldes rdau:P60493, die /{rR}oman/ enthalten

Bezeichnung	Häufigkeit
Roman	120.985
Kriminalroman	8.657
[Roman]	2.671
roman	2.657
Science-fiction-Roman	986
historischer Roman	937
Kriminal-Roman	903
Western-Roman	760
Roman aus d. amerikan. Westen	408
heiterer Roman	405

Da wir unsere Untersuchung auf Romane beschränken wollten, war für uns außerdem das Datenfeld ‚rdau:P60493‘ wichtig, das für Titelzusätze vorgesehen ist.⁵⁶ Der *DNB*-Katalog verzeichnet in diesem Feld relativ systematisch die Untertitel von Werken bzw. Büchern.⁵⁷ Bei den folgenden Statistiken ist also zu beachten, dass wir dadurch zwar den Großteil, aber keineswegs alle Werke abdecken, die man als ‚Romane‘ bezeichnen könnte, sondern nur diejenigen, die gemäß der Katalogisierungspraxis der *DNB* einen entsprechenden Eintrag im richtigen Feld bekommen haben (etwa weil diese Information auf dem Titelblatt eines katalogisierten Buches enthalten war). Tab. 1 zeigt die häufigsten Bezeichnungen in diesem Feld, die auf den regulären Ausdruck /{rR}oman/ matchen.⁵⁸ Insgesamt betrifft dies ca. 353.498 Bücher im Katalog⁵⁹ (wohl-gemerkt: Bücher, nicht Werke, denn darunter befinden sich gemäß dem Sammel-auftrag der *DNB* auch Neuauflagen, Übersetzungen usw.).

⁵⁶ Siehe <http://www.rdaregistry.info/Elements/u/#P60493> (letzter Aufruf 6.5.2018).

⁵⁷ Theoretisch hätten wir unseren Datensatz auch über die Sachgruppen zuschneiden können (etwa „08a Schöne Literatur“ oder „59 Belletristik“), allerdings ist die „Ausgabe der Sachgruppen im Linked-Data-Service [momentan] ausgesetzt worden“ (Datendienst *Bibliografische Dienstleistungen, Änderungen im Format RDF ab 12. September 2017 (Export-Release 03/2017)* (11. Juli 2017), 4, http://www.dnb.de/SharedDocs/Downloads/DE/DNB/service/rundschreiben20170711AenderungenRDF.pdf?__blob=publicationFile (letzter Aufruf 6.5.2018).

⁵⁸ Insgesamt gibt es 18.038 verschiedene /{rR}oman/-Bezeichnungen, von denen die meisten aber nur genau einmal vorkommen. Einige wenige false positives wie „romantische Oper“ fallen daher bei dieser Größenordnung nicht ins Gewicht, vgl. auch Tab. 1.

⁵⁹ Diese Zahl zeigt schon in Richtung der Größenordnung, die Gregory Crane im Jahr 2006 ins Spiel brachte, als er die Frage stellte: „What Do You Do with a Million Books?“ (in: *D-Lib Magazine* 12/3 [März 2006], <http://www.dlib.org/dlib/march06/crane/03crane.html> [letzter Aufruf 6.5.2018]). Der Unterschied besteht darin, dass Cranes Frage auf Textmining abzielte, wohingegen wir mit diesen Romanen nur auf Grundlage ihrer bibliographischen Metadaten rechnen.

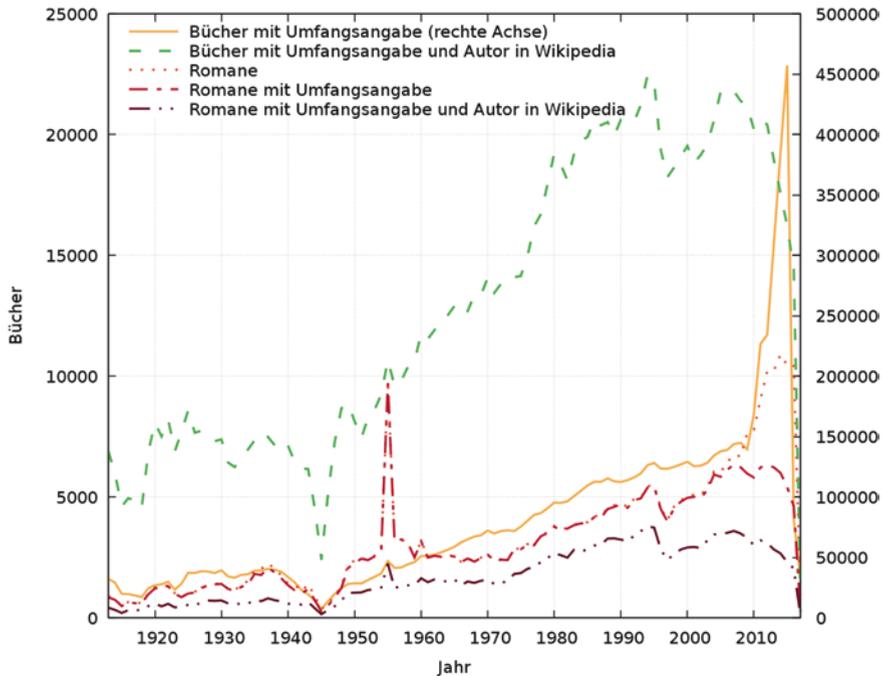


Abb. 2 Fünf verschieden qualifizierte Subdatensätze des *DNB*-Katalogs in zeitlicher Verteilung

Angesichts der Tatsache, dass Self-Publishing-Verlage wie der *GRIN Verlag* (mit insg. 121.456 Titeln) und *Books on Demand* (97.716 Titel) schon jetzt, nur wenige Jahre nach ihrer Gründung, den *DNB*-Katalog überflutet haben, setzten wir zusätzlich auf ein Relevanzkriterium, um unseren Datensatz weiter zu qualifizieren. Dafür haben wir auf *Wikidata* als Wissensdatenbank zurückgegriffen. Als relevant für diese Studie haben wir Autoren erachtet, die in mindestens einer *Wikipedia*-Sprachversion mit einem Artikel vertreten sind, die also die dortigen Relevanzhürden genommen haben (die Relevanzkriterien der verschiedenen *Wikipedias* sind teilweise problematisch, aber es gibt sie).⁶⁰

Abb. 2 zeigt die zeitliche Verteilung einiger Subdatensätze des *DNB*-Katalogs. Von den 14.102.309 Items im originalen *DNB*-Datensatz weisen – von den seit eigentlichem Sammlungsbeginn im Jahr 1913 erschienenen Werken – insgesamt 8.346.148 extrahierbare Seitenanzahlen auf (das sind 59 %) – die Gesamtsumme

⁶⁰Zum Filtern von Artikeln über literarische Autoren in der *Wikipedia* nach verschiedenen Relevanzkriterien vgl. auch Christoph Hube/Frank Fischer/Robert Jäschke/Gerhard Lauer/Mads Rosendahl Thomsen, *World Literature According to Wikipedia*, in: *arXiv:1701.00991 [cs.IR]* (4. Januar 2017), <https://arxiv.org/abs/1701.00991> (letzter Aufruf 6.5.2018); dort allerdings mit *DBpedia* statt *Wikidata* als Wissensdatenbank.

aller so extrahierbaren Seitenanzahlen beläuft sich momentan auf 1.327.973.922.⁶¹ Wenn wir die Gesamtmenge an Büchern auf ‚Romane‘ beschränken, kommen wir auf die genannten 353.498, von denen 316.518 auch Umfangangaben aufweisen und von denen wiederum 180.219 einen Verfasser haben, der mindestens einen *Wikipedia*-Eintrag hat. Diese letzte Zahl ist die Größe unseres Datensets, mit dem wir weiterarbeiten.

Der Code und die Dokumentation zum gesamten Operationalisierungsprozess liegt in unserem *GitHub*-Repositoryum.⁶² Die vorgestellte technische Lösung ist exemplarisch angelegt.⁶³ Unser Vorgehen sollte mit den Katalogdaten anderer Bibliotheken oder Institutionen sowie mit anderen Filtermechanismen reproduzierbar sein und *eine* mögliche empirische Grundlage für den gerade emergierenden Diskurs um eine allgemeine Textumfangstheorie schaffen.

4 Auswertung

4.1 Erster Blick in die *DNB*

Beginnen wir mit einer Übersicht, die uns etwas zur Repräsentativität der *DNB* sagt. Wer sind eigentlich die Autoren mit den meisten Büchern in der *Nationalbibliothek* (wie gesagt, inklusive Neuauflagen und Übersetzungen)? Eine Übersicht dazu gibt Tab. 2, hier noch bezogen auf alle Bücher, nicht eingeschränkt auf Romane. Goethe, Steiner, Hesse, Kosalik, Thomas Mann – so sieht die nach Bücheranzahl quantifizierbare Realität der *DNB* aus. Unter den 50 Schriftstellern mit den meisten Werken finden sich sechs Frauen. Die meistübersetzten fremdsprachigen Autoren in der *DNB* sind Shakespeare, Agatha Christie, Enyd Blythton, Lenin, Dostojewski, Tolstoi, Balzac, Jack London und Edgar Wallace.

Wenn wir diese Übersicht auf die Autorschaft von Romanen einschränken, ergibt sich ein etwas anderes Bild (Tab. 3, wiederum inklusive Neuauflagen und Übersetzungen). Vor allem (aber natürlich nicht nur) die Autoren von Roman-Bestsellern im U-Bereich treten hier hervor.

Schauen wir uns nun die Verteilung der Seitenanzahlen für die knapp 180.000 als ‚Roman‘ rubrizierten Bücher an (s. Abb. 3). Die Mehrzahl der Romane hat einen Umfang zwischen ca. 160 und 480 Seiten. Danach fällt die Kurve flacher, aber stetig weiter ab, bis wir es ab einer Seitenanzahl von ca. 1.600 nur mehr mit Ausreißern zu tun haben.

⁶¹ Da nicht alle Buchumfänge extrahierbar waren bzw. im Katalog verzeichnet sind, ist diese Zahl mit etwas Vorsicht zu genießen; sie gibt aber eine ganz gute Näherung ab (mehr als 1,3 Mrd. angesammelte Buchseiten in 105 Jahren *DNB*).

⁶² <https://github.com/weltliteratur/dnb/> (letzter Aufruf 6.5.2018).

⁶³ Vgl. auch Frank Fischer/Robert Jäschke, „Liebe und Tod in der Deutschen Nationalbibliothek. Der *DNB*-Katalog als Forschungsobjekt der digitalen Literaturwissenschaft“, in: *Konferenzabstracts zur DHd2018 in Köln* (26.2.–2.3.2018), 261–266, <https://doi.org/10.18716/KUPS.8085>.

Tab. 2 Autoren geordnet nach Anzahl der Werke im *DNB*-Katalog

Autor	Bücher
Johann Wolfgang von Goethe	5.160
Rudolf Steiner	3.992
Hermann Hesse	3.363
Heinz G. Konsalik	3.190
Thomas Mann	2.609
Karl Marx	2.577
Friedrich Engels	2.568
Stefan Zweig	2.532
Friedrich Schiller	2.047
Franz Kafka	2.020
Jacob Grimm	2.020
Theodor Storm	1.967
Bertolt Brecht	1.948
Wilhelm Grimm	1.943
Erich Kästner	1.873
Heinrich Böll	1.796
Friedrich Nietzsche	1.788
William Shakespeare	1.700
Rainer Maria Rilke	1.689
Marie Louise Fischer	1.678
Gottfried Keller	1.666
Anselm Grün	1.605
Martin Luther	1.590
Sigmund Freud	1.568
Theodor Fontane	1.549
E. T. A. Hoffmann	1.533
Agatha Christie	1.531
Karl May	1.503
Heinrich Heine	1.388
Christine Nöstlinger	1.344
Hedwig Courths-Mahler	1.302
Wilhelm Hauff	1.290
Immanuel Kant	1.280
Gert Fritz Unger	1.264
Heinrich von Kleist	1.250
Janosch	1.210

(Fortsetzung)

Tab. 2 (Fortsetzung)

Autor	Bücher
Enid Blyton	1.210
Wilhelm Busch	1.191
Adalbert Stifter	1.184
Erich Maria Remarque	1.181
Wladimir Iljitsch Lenin	1.117
Johanna Spyri	1.106
Fjodor Dostojewski	1.088
Lew Tolstoi	1.082
Honoré de Balzac	1.081
Max Frisch	1.079
Jack London	1.071
Lion Feuchtwanger	1.048
Edgar Wallace	1.041
Carl Jung	996

Beim Zoom auf einen Detailbereich des Diagramms, den Bereich von 200 bis 400 Seiten, fällt noch etwas anderes auf (s. Abb. 4). Und zwar gibt es einen regelmäßigen Peak der Seitenanzahlen aller 16 Seiten, was leicht zu erklären ist: 16 Seiten entsprechen genau einem Bogen, und es gibt nachvollziehbarerweise die Tendenz der Verleger und/oder Autoren, den Platz, den der letzte Bogen bietet, möglichst voll auszunutzen. Sowohl in Abb. 3 als auch in Abb. 4 zeigt sich übrigens, dass jeder zweite Peak höher als der jeweils vorhergehende ist, was auf den teilweisen Einsatz größerer Druckbogen hinweist. Diese hier unerwartet sichtbar werdenden buchhistorischen Fun-Facts sind immerhin dazu geeignet, genug Vertrauen in die Verlässlichkeit und Konsistenz der für uns wichtigen Metadaten zu etablieren, auf deren Basis die folgenden Auswertungen stattfinden.

4.2 Die umfangreichsten Bücher pro Verlag

In Abb. 5 wird die Anzahl von Romanen pro Verlag korreliert mit der mittleren Seitenanzahl pro Roman. Hier wird in der Tendenz deutlich, dass größere Verlage im Durchschnitt umfangreichere Bücher verlegen.

Um etwas genauer in die Verlagslandschaft zu schauen, haben wir in Tab. 4 aus den von uns gefilterten Katalogdaten die 20 Verlage mit den meisten Romanen bzw. Romanseiten in der *DNB* aufgelistet. Das gleichzeitige Vorkommen von „Rowohlt“ und „Rowohlt-Taschenbuch-Verl.“ sowie von abgekürzt „Fischer-Taschenbuch-Verl.“ und ausgeschrieben „Fischer-Taschenbuch-Verlag“ weist darauf hin, dass wir für diese Übersicht Verlagsnamen zunächst noch nicht normalisiert und zusammengeführt haben.

Tab. 3 Romanautoren
geordnet nach Anzahl der
Romane im *DNB*-Katalog

Autor	Bücher
Heinz G. Konsalik	2.232
Marie Louise Fischer	1.264
Gert Fritz Unger	1.013
Georges Simenon	783
Utta Danella	778
Edgar Wallace	654
Hedwig Courths-Mahler	647
Eleanor Hibbert	635
Pearl S. Buck	596
Alistair MacLean	582
Stephen King	577
Georgette Heyer	576
Agatha Christie	574
Theodor Fontane	565
Hans Ernst	563
Lion Feuchtwanger	501
Erich Maria Remarque	419
Hans Hellmut Kirst	411
Johannes Mario Simmel	403
Hans Fallada	396
Heinrich Mann	394
Fjodor Dostojewski	390
Barbara Cartland	390
Nora Roberts	381
Graham Greene	375
A. J. Cronin	370
Vicki Baum	366
Thomas Mann	359
Robert Ludlum	358
Gerd Hafner	357
Dean Koontz	354
Heinrich Böll	340
Alexandra Cordes	325
John le Carré	322
Marion Zimmer Bradley	321
Jason Dark	317

(Fortsetzung)

Tab. 3 (Fortsetzung)

Autor	Bücher
Willi Heinrich	313
Ludwig Ganghofer	311
Jack London	309
Joseph Roth	307
Danielle Steel	299
Johanna Lindsey	288
Erle Stanley Gardner	287
Siegfried Lenz	279
Jules Verne	277
Rosamunde Pilcher	274
Franz Kafka	271
Ernest Hemingway	271
Taylor Caldwell	269
Dorothy L. Sayers	269

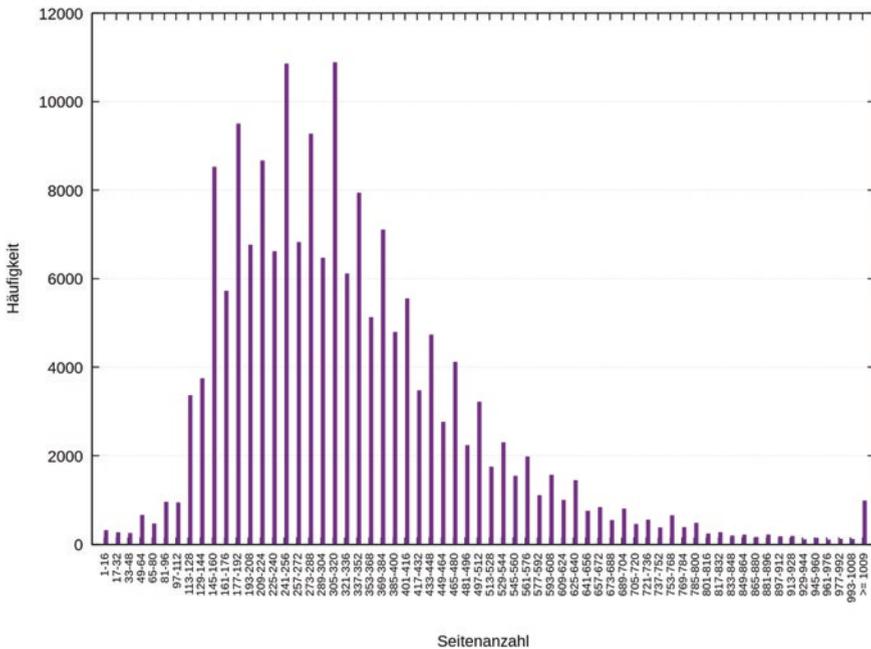


Abb. 3 Verteilung der Seitenanzahlen von 180.219 gefilterten Romanen aus dem DNB-Katalog

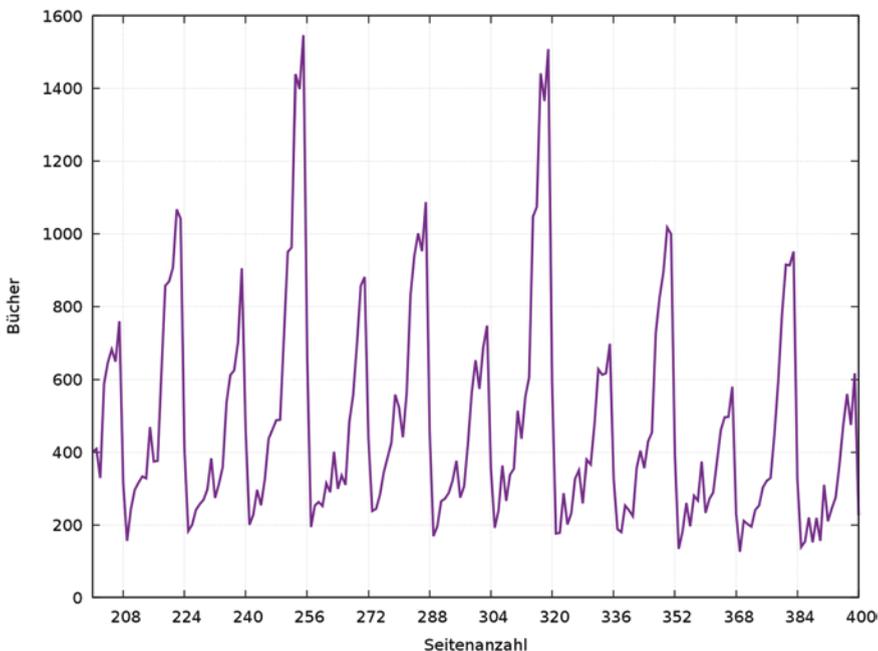


Abb. 4 Detailansicht des Bereichs 200–400 Seiten

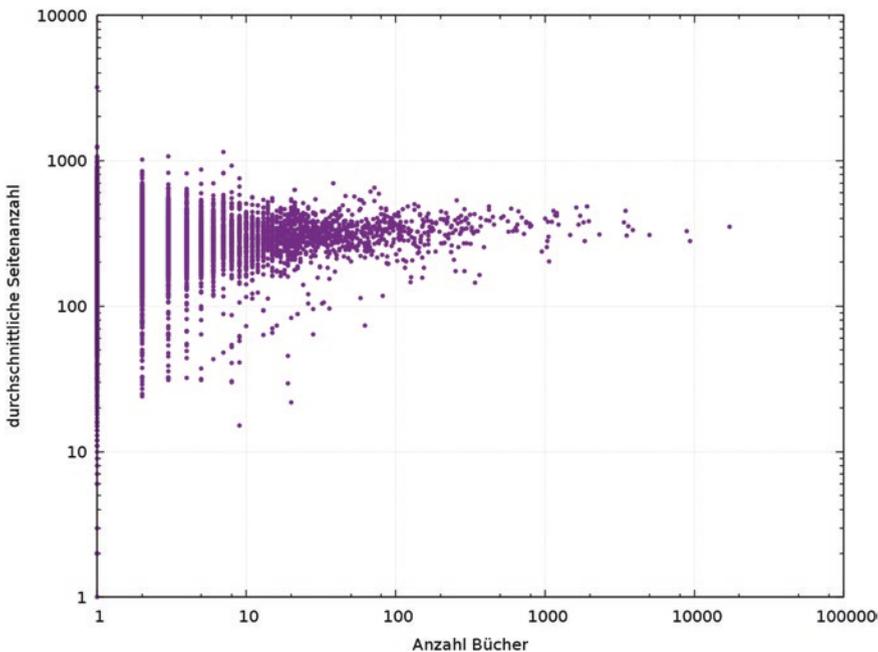


Abb. 5 Anzahl der Romane pro Verlag, korreliert mit der mittleren Seitenanzahl pro Roman (jeder Punkt entspricht einem Verlag)

Tab. 4 Top-20-Verlage nach Gesamtseitenanzahl aller als Roman markierten Katalog-Items

Verlag	Gesamt-Seitenanzahl	Bücher	Mittlere Seitenanzahl
Heyne	5.963.493	16.969	351
Goldmann	2.834.238	8.638	328
Rowohlt	2.561.798	9.195	278
RM-Buch-und-Medien-Vertrieb [u. a.]	1.517.759	3.355	452
Ullstein	1.499.865	4.872	307
Fischer-Taschenbuch-Verl	1.258.903	3.535	356
Dt. Taschenbuch-Verl	1.247.960	3.743	333
Piper	1.224.204	3.237	378
Suhrkamp	1.058.664	3.467	305
Weltbild	896.222	1.857	482
Dt. Buch-Gemeinschaft	735.110	1.920	382
Diogenes	711.362	2.290	310
Blanvalet	704.751	1.483	475
Droemer Knauer	689.177	1.653	416
Büchergilde Gutenberg	677.264	1.800	376
Rowohlt-Taschenbuch-Verl	596.779	1.640	363
Aufbau-Verl	517.749	1.186	436
Fischer-Taschenbuch-Verlag	516.420	1.844	280
Dt. Bücherbund	507.782	1.123	452
Lübbe	486.821	1.159	420

Dies geschieht allerdings in Tab. 5, in der 25 ausgewählte Literaturverlage nach absteigender durchschnittlicher Seitenanzahl gerankt sind. Die verschiedenen Schreibungen von Verlagsnamen wurden hierfür zusammengeführt, was die Unterschiede zur vorherigen Tabelle erklärt.

Romane, die im *Manesse Verlag* erscheinen, der sich die „ambitionierte Klassikervermittlung“ auf die Fahnen geschrieben hat, sind im Mittel am umfangreichsten. Wie in Abb. 5 bereits beobachtet, scheinen größere Verlage nicht nur mehr, sondern auch längere Romane zu verlegen. Diese Zahlen sollten freilich nicht losgelöst von anderen Faktoren betrachtet werden.

Sehen wir noch einmal von der Verlagskategorie ab und betrachten den gesamten Datensatz: Abb. 6 zeigt, wie sich der mittlere Buchumfang der Romane zwischen den Jahren 1913 und 2017 entwickelt hat. Dieses Bild könnte man nun z. B. mit einer Vermutung des Verlegers Helge Malchow konfrontieren: „Mit dem Einzug des Computers in den Schreibprozess sind die Romane nach meinem

Tab. 5 Daten zu 25 ausgewählten Literaturverlagen, sortiert nach mittlerer Seitenanzahl

Verlag	Gesamt-Seitenanzahl	Bücher	Mittlere Seitenanzahl
Manesse	202.995	381	532
Rogner & Bernhard	8.878	21	422
Aufbau	1.178.750	2.885	408
Hoffmann und Campe	369.031	967	381
Piper	1.231.254	3.257	378
Goldmann	3.195.060	8.675	368
Insel	373.911	1.028	363
Heyne	6.041.445	17.138	352
Hanser	291.366	831	350
Luchterhand Literaturverlag	254.858	764	333
Verbrecher Verlag	12.290	37	332
Kiepenheuer & Witsch	415.369	1.261	329
Schöffling & Co	46.841	143	327
Eichborn	111.541	341	327
Suhrkamp	1.073.405	3.507	306
Berlin Verlag	67.871	223	304
Ullstein	166.7197	5.470	304
Ammann	44.341	146	303
Rowohlt	3.258.380	11.140	292
Voland & Quist	2.349	9	261
Reclam	278.548	1.072	259
Wiesenburg	4.251	17	250
Blumenbar	7.166	29	247
Wallstein	13.587	57	238
Nagel & Kimche	51.367	223	230

Eindruck im Durchschnitt um zehn Prozent länger geworden.“⁶⁴ Die Daten scheinen diese Verlegerempfindung zu bestätigen, zumindest für den Zeitraum zwischen Anfang der 1980er Jahre (mit den ersten Heimcomputern) und etwa 2010.

⁶⁴ Juli Zeh, „Wie wollen wir lesen?“ [Werkstattgespräch mit den Verlegern Michael Krüger und Helge Malchow], in: *ZEIT ONLINE* (15. November 2012), <http://www.zeit.de/2012/47/Buecher-Internet-Literatur-Verleger>.

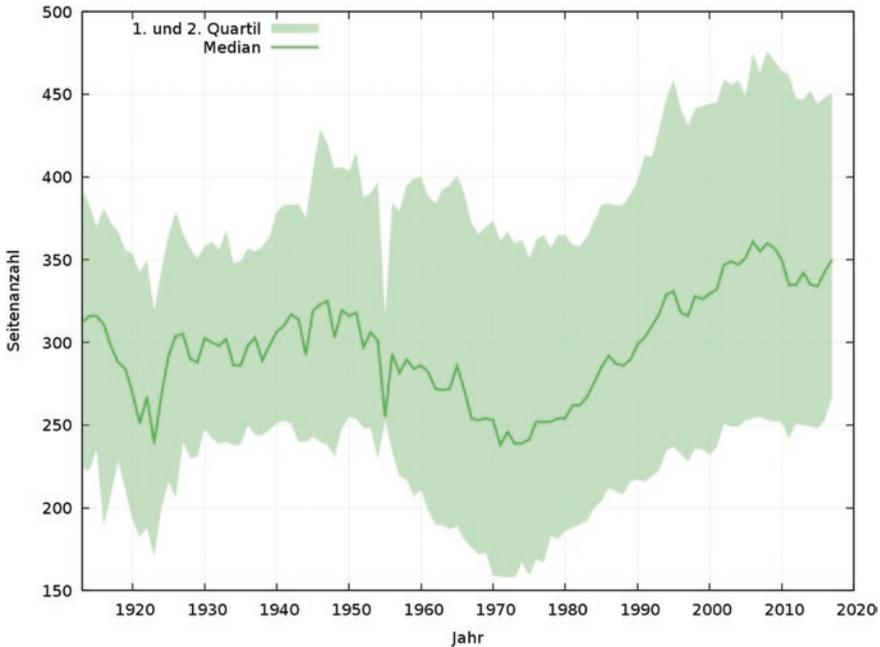


Abb. 6 Entwicklung der mittleren Seitenanzahl pro Jahr seit 1913

Da sich die Diskussion um Textumfänge gut anhand von Extremtexten führen lässt, interessierte uns als nächstes, welche Autoren und Romane die Verlage als ihre umfangreichsten platziert haben. Wie bereits bemerkt, steigert das Operieren im 1.000-Seiten-Buch-Segment das ökonomische Risiko, und nicht jeder Verlag kann sich dies leisten (von den oben ausgewählten Literaturverlagen haben kleinere wie etwa *Blumenbar*, *Nagel & Kimche*, *Voland & Quist* und *Wiesenburg* bisher keine Tausendseiter in ihrer Backlist).

Auch bei den größeren Verlagen haben wir es meist mit nur einer Handvoll Tausendseitern zu tun, an denen sich zumindest Teile der Verlagspolitik ablesen lassen:

An der Liste zum *Ammann Verlag* (Tab. 6) fällt zuerst auf, dass zwei Bücher fehlen: Dostojewskis *Die Brüder Karamasow* (2003, 1.271 Seiten)⁶⁵ und *Ein grüner Junge* (2006, 829 Seiten)⁶⁶ – dies auch noch einmal als Hinweis darauf, dass wir bei der Analyse auf die Katalogdaten angewiesen sind und Wege finden müssen, diese zu verifizieren. Jedenfalls handelt es sich bei all diesen Dostojewski-Ausgaben um die Neuübersetzungen Swetlana Geiers, und es wird anhand

⁶⁵ Siehe <http://d-nb.info/96916162X> (letzter Aufruf 6.5.2018).

⁶⁶ Siehe <http://d-nb.info/1006191569> (letzter Aufruf 6.5.2018).

Tab. 6 Umfangreichste Romane des Ammann Verlags

Seitenanzahl	Autor: Titel (Erscheinungsjahr der Ausgabe)
962	Fjodor Dostojewski: Böse Geister (1998)
909	Fjodor Dostojewski: Der Idiot (1996)
766	Fjodor Dostojewski: Verbrechen und Strafe (1994)
607	Svend Åge Madsen: Sieben Generationen Wahnsinn (2000)
572	Ulrich Peltzer: Stefan Martinez (1995)

Tab. 7 Umfangreichste Romane des *Manesse Verlags*

Seitenanzahl	Autor: Titel (Erscheinungsjahr der Ausgabe)
1.347	Victor Hugo: Die Elenden (1968)
1.236	Fjodor Dostojewski: Die Brüder Karamasow (1964)
1.196	Charles Dickens: David Copperfield (1961)
1.193	Thomas Wolfe: Von Zeit und Fluss (2014)
1.155	William Makepeace Thackeray: Jahrmarkt der Eitelkeit (1959)
1.145	George Eliot: Middlemarch (1962)
1.062	George Eliot: Daniel Deronda (1994)
990	Stefan Żeromski: In Schutt und Asche (1988)
956	Charlotte Brontë: Shirley (1989)
953	Fjodor Dostojewski: Schuld und Sühne (1985)

dieses Rankings einmal konkret bezifferbar, warum Geier „Die Frau mit den 5 Elefanten“ genannt wurde (vgl. auch den gleichnamigen Dokumentarfilm über sie, Regie: Vadim Jendreyko, veröffentlicht 2009). Ihre Neuübersetzungen der fünf großen Dostojewski-Elefanten waren ein Insignium des mittlerweile aufgelösten Schweizer Verlags.

Wie schon oben anhand der durchschnittlichen Seitenanzahlen deutlich wurde: Schwerpunkt von *Manesse* sind schwergewichtige Klassiker (Tab. 7).

Die nach der Seitenanzahl umfangreichsten im *DNB*-Katalog verzeichneten Items von *Suhrkamp* sind keine Einzelwerke, sondern die großformatigen, broschiierten Bände der *Suhrkamp Quarto*-Reihe (etwa Amos Oz: *Die Romane* [2009], 2.569 Seiten)⁶⁷. Durch unsere Beschränkung auf Romane ergibt sich eine sinnvollere Übersicht über das Tausendseiter-Segment von *Suhrkamp*, meistens kanonisierte Romane (Tab. 8).

Auch bei *Rowohlt* finden sich in diesem Segment neben Romanen unserer unmittelbaren Gegenwart viele seit langem kanonisierte Werke, etwa Musils *Mann ohne Eigenschaften* (Tab. 9).

⁶⁷ Siehe <http://d-nb.info/991420225> (letzter Aufruf 6.5.2018).

Tab. 8 Umfangreichste Romane des *Suhrkamp Verlags*

Seitenanzahl	Autor: Titel (Erscheinungsjahr der Ausgabe)
1.195	Peter Weiss: Die Ästhetik des Widerstands (2005)
1.162	Amos Oz: Eine Geschichte von Liebe und Finsternis (2016)
1.122	James Joyce: Ulysses (2004)
1.025	William T. Vollmann: Europe Central (2013)
1.018	Clemens J. Setz: Die Stunde zwischen Frau und Gitarre (2017)
972	Ippolito Nievo: Pisana oder Die Bekenntnisse eines Achtzigjährigen (1956)
972	Uwe Tellkamp: Der Turm (2008)
941	Jaume Cabré: Die Stimmen des Flusses (2016)
923	Elsa Morante: Lüge und Zauberei (1981)
891	Martin Walser: Halbzeit (1990)

Tab. 9 Umfangreichste Romane des *Rowohlt Verlags*

Seitenanzahl	Autor: Titel (Erscheinungsjahr der Ausgabe)
1.723	Péter Nádas: Parallelgeschichten (2013)
1.671	Robert Musil: Der Mann ohne Eigenschaften (1952)
1.595	Thomas Pynchon: Gegen den Tag (2008)
1.342	Harold Brodkey: Die flüchtige Seele (1995)
1.303	Péter Nádas: Buch der Erinnerung (1992)
1.230	Irving Stone: Der Seele dunkle Pfade (2001)
1.228	Navid Kermani: Dein Name (2015)
1.210	Rosamunde Pilcher: Heimkehr (2014)
1.196	Thomas Wolfe: Von Zeit und Strom (1989)
1.193	Thomas Pynchon: Die Enden der Parabel (1989)

Der umfangreichste Roman von *Kiepenheuer & Witsch* (Tab. 10), die deutsche Übersetzung von David Foster Wallaces *Unendlicher Spaß* (erschieden 2009, 1.547 Seiten)⁶⁸ fehlt in der Übersicht, da er im entsprechenden Datenfeld nicht als ‚Roman‘ verschlagwortet ist (das Filtern der für den Roman vergebenen Sachgruppe „B Belletristik“ hätte hier Abhilfe schaffen können, allerdings ist die Aufnahme der Sachgruppen, wie oben berichtet, im Datenabzug der *DNB* momentan ausgesetzt).

Dabei lässt sich an diesem Beispiel zeigen, dass ein mehr als tausendseitiges Buch immer ein Ereignis ist: Das Buch war so dick, dass es bei der Herstellung

⁶⁸ <http://d-nb.info/994243626> (letzter Aufruf der Links 6.5.2018).

Tab. 10 Umfangreichste Romane bei *Kiepenheuer & Witsch*

Seitenanzahl	Autor: Titel (Erscheinungsjahr der Ausgabe)
1.320	Frank Schätzing: Limit (2009)
1.035	Manès Sperber: Wie eine Träne im Ozean (1961)
1.033	Erwin Strittmatter: Der Laden (1989)
1.001	René Schickele: Das Erbe am Rhein (1965)
997	Frank Schätzing: Der Schwarm (2004)
965	Don DeLillo: Unterwelt (1998)
964	Frank Schätzing: Breaking News (2014)
892	Maxim Biller: Biografie (2016)
874	André Brink: Zeit des Terrors (1994)
861	Holger Karsten Schmidt: Isenhart (2012)

nur gerade noch so durch die Bindestraße passte, was mitunter zu Herstellungsfehlern führte.⁶⁹ Eine sicher nicht neutrale, aber in ihrer Pointierung gut in unsere Argumentation passende Einschätzung der dedizierten Bewerbung dieses Buch-Events stammt von Rainald Goetz, geäußert bei seinem Auftritt in der *Harald Schmidt Show* am 8. April 2010: „Wir wollen Helge Malchow, den Chef des Kiepenheuer-Verlages, preisen, [...], der hat die Leute so bequatscht, dass alle Angst hatten irgendwie, dieses sehr schlechte Buch schlecht zu finden.“⁷⁰ Das Buch wurde jedenfalls ein Bestseller, und die zur Schau gestellte Dickleibigkeit trug sicher dazu bei.

Der kleine Streifzug durch den Elefantenbestand einiger Literaturverlage hat gezeigt, dass man nur anhand der umfangreichsten Werke, die der Verlagskatalog zu bieten hat, Teile der Verlags-DNA ableiten kann.⁷¹ Das Verlegen dicker Bücher ist immer auch eine Setzung. Ein hilfreicher nächster Schritt zur Theoriebildung wäre es, einige Vertreter aus der Liga der Tausendseitigkeit einer Fallstudie zu unterziehen, um den Zusammenhang zwischen Buchumfang und Rezeption einmal konkret anhand einiger Beispiele zu untersuchen und der quantitativen eine qualitative Perspektive zur Seite zu stellen.

⁶⁹Vgl. Konrad Geyer [d. i. Wolf Schmid], „Die Unendlichkeit reklamiert“, in: *Kommentarblog* (22. Januar 2013), <https://kommentarblog.wordpress.com/2013/01/22/die-unendlichkeit-reklamieren/>; dabei gibt es weitaus dickere Bücher, Rekordhalter ist eine 322 mm dicke Ausgabe mit allen *Miss Marple-Stories* (Stand 2009, vgl. o.N., „Thickest book published“, in: *Guinness World Records*, <http://www.guinnessworldrecords.com/world-records/thickest-book-published> [letzter Aufruf 6.5.2018]).

⁷⁰Rainald Goetz/Harald Schmidt, [„Interview“], in: *Harald Schmidt Show* (ARD, 8. April 2010), inoffizieller Link zum Video: <https://www.youtube.com/watch?v=BqDv6F9eTHA> (letzter Aufruf 6.5.2018).

⁷¹In unserem GitHub-Repository finden sich derlei Auflistungen zu weiteren Verlagen, <https://github.com/weltliteratur/dnb/> (letzter Aufruf 6.5.2018).

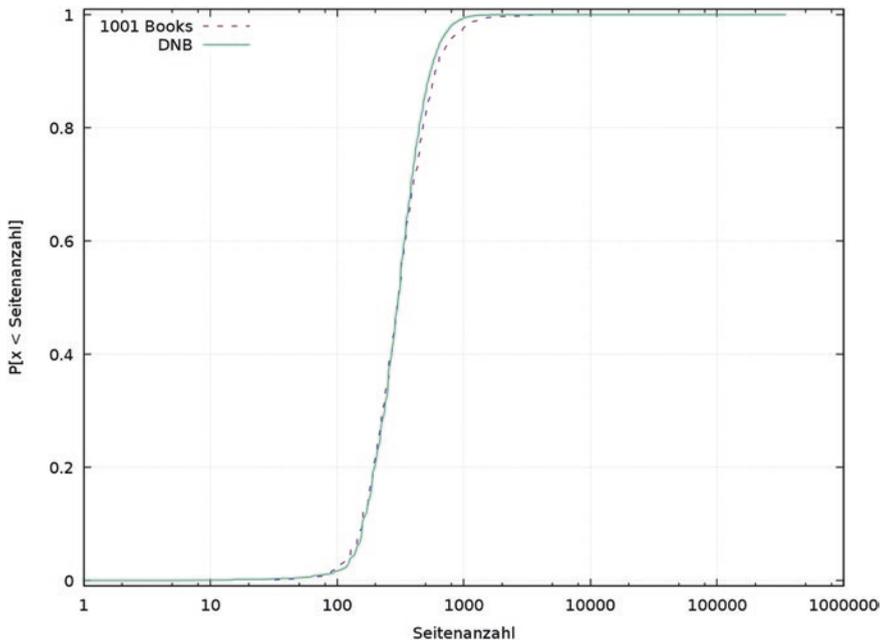


Abb. 7 Vergleich der Verteilung der Seitenanzahlen zwischen dem *DNB*-Katalog und dem *1001-Books*-Kanon

4.3 Textumfang und Kanonisierung

Eine oben gestellte Frage war die nach dem Verhältnis zwischen Textumfang und Kanon. Abb. 7 visualisiert die kumulierte Verteilungsfunktion der Seitenanzahlen unseres *DNB*-Datensatzes sowie die des *1001-Books*-Kanon. Die Vergleichbarkeit beider Datensätze sollte dadurch gegeben sein, dass in Boxalls Kanon „das Wort ‚Buch‘ kurzerhand mit ‚Roman‘ gleichgesetzt wird“, ⁷² wir also letztlich zwei Romankorpora vergleichen. Dass wir es bei den *1001 Books* mit Umfangangaben zu englischsprachigen Romanen (bzw. Übersetzungen) zu tun haben und im Fall unseres *DNB*-Datensatzes mit Umfangsinformationen zu deutschsprachigen Romanen (bzw. Übersetzungen), ist natürlich ein hier nicht eingerechneter Faktor. Wir betrachten Seitenanzahlen für diesen Vergleich also als trockenes, von anderen Faktoren unbeeinflusstes Faktum (was sie natürlich, siehe oben, nicht sind).

Wenn wir uns also ganz auf die Größe der Buchblöcke konzentrieren, dann sagt der Verlauf der beiden Kurven etwas über die Wahrscheinlichkeit aus, dass ein Buch des jeweiligen Datensatzes eine gegebene Anzahl von Seiten unterschreitet.

⁷²Dorowin (Anm. 47).

Der spätere Anstieg der Kurve des *1001-Books*-Kanons impliziert, dass in diesem Datensatz der Anteil der umfangreicheren Bücher größer ist als im *DNB*-Datensatz. Die Bücher in letzterem zeigen also eine deutliche Tendenz, umfangreicher zu sein.

Im Bereich 100–400 Seiten befinden sich 68,1 % des 1.001er-Kanons und 71,7 % der *DNB*-Romane: vergleichbare Zahlen. Wenn wir allerdings den Bereich ab 1.000 Seiten betrachten, ändert sich das Bild: Von den Büchern im Kanon überschreiten 2,3 % diese Seitengrenze, von den *DNB*-Romanen nur 0,6 %. Im *1001-Books*-Kanon befinden sich also überdurchschnittlich viele umfangreiche Romane. An diesem Beispiel zeigt sich, dass Texte extremen Umfangs jenseits der 1.000 Seiten tatsächlich kanonfähiger sein können, ein erster datengestützter Hinweis auf den Zusammenhang zwischen Textumfang und Kanonisierung, den es freilich umfassender zu untersuchen gilt.

5 Ausblick

Die hier angestellten empirischen Versuche vermessen Texte in ihrer Buch gewordenen Gestalt. Unser empirischer Ansatz ist seitenbasiert, nicht wörter- oder zeichenbasiert. Die beiden letzteren Maße sollten aber langfristig ins Metadatenarsenal literarischer Werke aufgenommen werden. Inwiefern Prozentangaben (etwa für den Umfang von Textteilen oder für die Position einer Textstelle) erkenntnisfördernd sind, wäre noch zu eruieren. Um komparatistische Perspektiven zu entwickeln, muss jedenfalls das Instrumentarium feststehen: „The concept of length is [...] fixed when the operations by which length is fixed are fixed“.⁷³

Langfristig wird es für eine Theorie des literarischen Textumfangs darauf ankommen, die eher materialen bzw. buchhistorischen und die eher digital-quantifizierenden Komponenten miteinander zu verbinden. Wir haben in diesem Aufsatz versucht, beide Diskurse anzustoßen, mit einem Bibliothekskatalog als einer möglichen Schnittstelle.

Danksagung Wir danken der *Information School* der *University of Sheffield* für die Bewilligung eines Visiting Scholarships im Mai 2017, das uns ermöglicht hat, gemeinsam an der Operationalisierung zu arbeiten.

Für Kritik und Anregungen, die ihren Niederschlag in diesem Aufsatz gefunden haben, danken wir Peggy Bockwinkel, J. Berenike Herrmann, Jörg Ossenkopp, Annika Rockenberger, Michael Rölcke, Joseph Wälzholz und Heiko Zimmermann. Wir danken außerdem Steffen Martus und Stefan Matuschek für die produktive Kritik der allerersten Projektskizze von Ende 2012 sowie dem Göttingen *Centre for Digital Humanities* für die Gelegenheit, das Projekt im Juni 2014 zum ersten Mal öffentlich vorstellen zu dürfen.

⁷³ P. W. Bridgman, *The Logic of Modern Physics*, New York 1927, 5.

Literatur⁷⁴

- Aira, César, *El Congreso de Literatura. Mérida: Fundación Casa de las Letras „Mariano Picón Salas“*, Mérida 1997.
- Aristoteles, *Poetik*, Griechisch/Deutsch, übers. und hg. von Manfred Fuhrmann, Stuttgart 1997 [1982].
- Baumanski [d. i. Fabian Baumann], „362.597 Seiten“, in: *Der Umblätterer* (20. August 2012), <http://www.umblaetterer.de/2012/08/20/362597-seiten/>.
- Bayard, Pierre, *Comment parler des livres que l'on n'a pas lus ?*, Paris 2007.
- Blumenberg, Hans, *Beschreibung des Menschen*, aus dem Nachlaß, hg. von Manfred Sommer, Frankfurt/M. 2006.
- Borges, Jorge Luis, *Ficciones – El Aleph – El Informe de Brodie*, Caracas 1993.
- Boxall, Peter (Hg.), *1001 Books You Must Read Before You Die*, London 2006.
- Bridgman, P. W., *The Logic of Modern Physics*, New York 1927.
- Crane, Gregory, „What Do You Do with a Million Books?“, in: *D-Lib Magazine*, 12/3 (März 2006), <http://www.dlib.org/dlib/march06/crane/03crane.html>.
- Datendienst *Bibliografische Dienstleistungen*, *DNBTitel.rdf.gz* (14504779 Datensätze/235578087 Tripel/DNB-Titel-Daten im Format RDF (RDF/XML), Stand: Juni 2017 (erzeugt am 23. Juni 2017), <http://www.dnb.de/datendienst>.
- Dorowin, Hermann, „Zu kurz für den Kanon? Zum Problem der ‚kleinen Form‘“, in: Sanna, Simonetta (Hg.), *Der Kanon in der deutschen Sprach- und Literaturwissenschaft*, Bern u. a. 2009, 137–144.
- Drożdż, Stanisław/Oświecimka, Paweł/Kulig, Andrzej u. a., „Quantifying origin and character of long-range correlations in narrative texts“, in: *Information Sciences* 331 (20. Februar 2016), 32–44.
- Enzensberger, Hans M., *Meine Lieblings-Flops, gefolgt von einem Ideen-Magazin*, Berlin 2011.
- Ercolino, Stefano, *The Maximalist Novel. From Thomas Pynchon's Gravity's Rainbow to Roberto Bolaño's 2666*, New York/London 2014.
- Escobar Ulloa, Ernesto, „Entrevista a César Aira“, in: *The Barcelona Review. Revista Internacional de Narrativa Breve Contemporánea*, 44 (Sept./Okt. 2004), http://www.barcelonareview.com/44/s_ca.htm.
- Fellinger, Raimund, „Schreiben: Sich zur Ruhe setzen“. Die Entstehung von Mein Jahr in der Niemandsbucht“, in: Klaus Kastberger (Hg.), *Peter Handke. Freiheit des Schreibens – Ordnung der Schrift*, Wien 2009, 133–142, 164–173, <http://handkeonline.onb.ac.at/forschung/pdf/fellinger-2009.pdf>.
- Fischer, Frank u. a., „Lob des 100-Seiten-Buchs“, in: *EDIT. Papier für neue Texte* 61 (Frühjahr 2013), 117–129.
- Fischer, Frank/Veentjer, Ubbo, „Empirical Data on Over-Length Books“, in: *weltliteratur.net* (25. Mai 2016), <http://weltliteratur.net/Empirical-Data-on-Over-Length-Books/>.
- Fischer, Frank/Jäschke, Robert, „Liebe und Tod in der Deutschen Nationalbibliothek. Der DNB-Katalog als Forschungsobjekt der digitalen Literaturwissenschaft“, in: *Konferenzabstracts zur DHd2018 in Köln* (26.2.–2.3.2018), 261–266, DOI <https://doi.org/10.18716/KUPS.8085>.
- Forster, E. M., *Aspects of the Novel*, hg. von Oliver Stallybrass, Harmondsworth 1976 [1927].
- Genette, Gérard, *Seuils*, Paris 1987.

⁷⁴ Alle angegebenen URLs wurden am 6.5.2018 auf Erreichbarkeit geprüft.

- Geyer, Konrad [d. i. Wolf Schmid], „Die Unendlichkeit reklamiert“, in: *Kommentarblog* (22. Januar 2013), <https://kommentarblog.wordpress.com/2013/01/22/die-unendlichkeit-reklamieren/>.
- Goetz, Rainald/Schmidt, Harald, [„Interview“], in: *Harald Schmidt Show* (ARD, 8. April 2010), inoffizieller Link zum Video: <https://www.youtube.com/watch?v=BqDv6F9eTHA>.
- Gumbrecht, Hans U., *Stimmungen lesen. Über eine verdeckte Wirklichkeit der Literatur*, München 2011.
- H., Anja, „Zeitbegrenzung zum Lesen eines E-Books?“, in: *Kindle Hilfeforum* (22. Dezember 2014), <https://www.amazon.de/gp/help/customer/forums/kindleqnal/?cdThread=Tx36D9B96YMG8S>.
- Handke, Peter, *Mein Jahr in der Niemandsbucht. Ein Märchen aus den neuen Zeiten*, Frankfurt a. M. 1994.
- Hänssler, Boris, „Wissenschaft ohne Geist“, in: *Süddeutsche Zeitung* (15. Mai 2013), 16.
- Häntzschel, Günter/Hummel, Adrian/Zedler, Jörg, *Deutschsprachige Buchkultur der 1950er Jahre. Fiktionale Literatur in Quellen, Analysen und Interpretationen*, Wiesbaden 2009.
- Heßelmann, Peter, „Herzog Anton Ulrichs Roman *Römische Octavia* in neuer Perspektive“, Rezension zu: Stephan Kraft: *Geschlossenheit und Offenheit der „Römischen Octavia“* von Herzog Anton Ulrich. „Der roman macht die ewigkeit gedencken, den er nimbt kein endt“, Würzburg 2004, in: *IASOnline* (5. November 2004), http://www.iasonline.de/index.php?vorgang_id=1029.
- Hube, Christoph/Fischer, Frank/Jäschke, Robert u. a., *World Literature According to Wikipedia*, in: *arXiv:1701.00991 [cs.IR]* (4. Januar 2017), <https://arxiv.org/abs/1701.00991>.
- Jostmann, Nils B./Lakens, Daniël/Schubert, Thomas W., „Weight as an Embodiment of Importance“, in: *Psychological Science* 20/9 (September 2009), 1169–1174, <http://www.jstor.org/stable/40575159>.
- Küveler, Jan, „Einmal quer durch Faserland“, in: *Welt am Sonntag* (20. November 2016), 57 f.
- Miller, Laura, „Are longer books more important?“, in: *Salon.com* (6. September 2012), http://www.salon.com/2012/09/05/are_longer_books_more_important/.
- Moretti, Franco, „Conjectures on World Literature“, in: Ders., *Distant Reading*, London/New York 2013 [2000], 43–62.
- Moretti, Franco, „Style, Inc. Reflections on Seven Thousand Titles (British Novels, 1740–1850)“, in: *Critical Inquiry* Vol. 36/1 (Herbst 2009), 134–158.
- O.N., „Longest novel“, in: *Guinness World Records*, <http://www.guinnessworldrecords.com/world-records/longest-novel>.
- O.N., „Notizen. Schriftstellerklagen aus den sechziger Jahren des vorigen Jahrhunderts“, in: *Blätter für literarische Unterhaltung* 27, Bd. 2: *Juli bis December* (10. September 1863), 682 f.
- O.N., „Thickest book published“, in: *Guinness World Records*, <http://www.guinnessworldrecords.com/world-records/thickest-book-published>.
- Paco [d. i. Frank Fischer], „Tausendseiter“, in: *Der Umblätterer* (21. Oktober 2014), <http://www.umblaetterer.de/2014/10/21/tausendseiter/>.
- Platthaus, Andreas, „Kindisch klingt der Ruf des Ku[c]kucks“, in: *Frankfurter Allgemeine Zeitung* (1. Dezember 2001), V.
- Raddatz, Fritz J., *Tagebücher 2002–2012*, Reinbek b. Hamburg 2014.
- Reich-Ranicki, Marcel, „Wer will, soll’s besser machen“, in: *Die Welt* (21. Oktober 2013), <https://www.welt.de/102852275>.
- Schlaffer, Heinz, „Der Umgang mit Literatur. Diesseits und jenseits der Lektüre“, in: *Poetica* 31 (1999), 1–25.
- Schmidt, Arno, „Ich bin erst sechzig“ [1955], in: *Bargfelder Ausgabe*, Werkgruppe I, Band 4. Zürich 1987.
- Schmidt, Arno, „*Fouqué*“-Typoskript, DLA Marbach, zit. n. Stephan Reuter, *Arno Schmidt und die Bücher – betrachtet aus der Perspektive einer gestrichenen Textstelle der Fouqué-Biographie*, Webpublikation [2000], ursprüngliche URL: <http://www.stephanreuthner.de/>

- [smbio7.htm](http://web.archive.org/web/20110415082919/www.stephanreuthner.de/smbio7.htm) (mittlerweile offline, Zugriff noch möglich über das *Internet Archive*: <http://web.archive.org/web/20110415082919/www.stephanreuthner.de/smbio7.htm>).
- Schneider, Jost, *Sozialgeschichte des Lesens. Zur historischen Entwicklung und sozialen Differenzierung der literarischen Kommunikation in Deutschland*, Berlin/New York 2004.
- Schöch, Christof, „Wiederholende Forschung in den digitalen Geisteswissenschaften“, in: *Konferenzabstracts zur DHd2017 in Bern* (13.–18.2.2017), 207–212, http://www.dhd2017.ch/wp-content/uploads/2017/02/Abstractband_ergaenz.pdf.
- Spoerhase, Carlos, *Linie, Fläche, Raum. Die drei Dimensionen des Buches in der Diskussion der Gegenwart und der Moderne* (Valéry, Benjamin, Moholy-Nagy), Göttingen 2016.
- Süwolto, Leonie, „Pragmatik und Ästhetik des Kleinen. Literarische, visuelle und mediale Mikroformate im 20. und 21. Jahrhundert“ [Konferenzankündigung], in: *H/Soz/Kult* (15. März 2013), <http://www.hsozkult.de/event/id/termine-21391>.
- Tellkamp, Uwe, *Der Turm. Geschichte aus einem versunkenen Land*, Frankfurt a. M. 2008.
- Thomas, Lindsay/Brennan, Shannon, *On Length* (Ausschreibung für ein geplantes, dann aber abgelehntes Seminar für die ACLA-Konferenz 2017) 2016, nur noch über das *Internet Archive* zugänglich: <https://web.archive.org/web/20160921040650/http://acla.org/length>.
- Tietze, Mark-Stefan, „Lesedauer: 100 Minuten“, in: *die tageszeitung* (13. Juni 2017), <http://www.taz.de/!5416422/>.
- UNESCO, *Recommendation concerning the International Standardization of Statistics Relating to Book Production and Periodicals*, Paris 1964, http://portal.unesco.org/en/ev.php-URL_ID=13068&URL_DO=DO_TOPIC&URL_SECTION=201.html.
- Vasilyev, Gennady, *Wiener Moderne. Diskurse und Rezeption in Russland*, Berlin 2015.
- Wälzholz, Joseph, „Clemens Brentano und das Geheimnis der Seitenzahlen“, in: *Die Welt* (13. September 2014), <https://www.welt.de/132206459>.
- Zaschke, Christian, „Klopfer für die Ewigkeit“, in: *Süddeutsche Zeitung* (16. Dezember 2009), 3.
- Zeh, Juli, „Wie wollen wir lesen?“ [Werkstattgespräch mit den Verlegern Michael Krüger und Helge Malchow], in: *ZEIT ONLINE* (15. November 2012), <http://www.zeit.de/2012/47/Buecher-Internet-Literatur-Verleger>.

Online-Ressourcen

- [Der Umblätterer] *100 Seiten. Ein Kanon kurzer Bücher*, <https://www.umblaetterer.de/100-seiten/>.
- GitHub-Arbeitsrepositorium*, <https://github.com/weltliteratur/dnb/>.
- RDA Registry*, <http://www.rdaregistry.info/Elements/u/#P60493>
- [Reclam] *100 Seiten für 100 Minuten*, <https://www.reclam.de/100Seiten>.
- Wikidata*, <https://www.wikidata.org/>.
- [Wikipedia,] *List of longest novels*, Stand vom 22.8.2017, https://en.wikipedia.org/wiki/List_of_longest_novels.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Reassembling the Novel. The English Novel, 1789–1919

Allen Riddell und Michael Betancourt

1 Introduction

An inclusive history of the novel during the long 19th century, one involving the demography of novels and novelists, has been widely discussed. A precondition for this kind of historical research is an exhaustive list of novels published during the period. The contribution of this paper are estimates of yearly rates of new novel production between 1837 and 1919 in the British Isles.¹ James Raven,

¹Following Peter Garside/James Raven/Rainer Schöwerling, “General Introduction,” in: Peter Garside, James Raven, Rainer Schöwerling (eds.), *The English Novel, 1770–1829: A Bibliographical Survey of Prose Fiction Published in the British Isles*, Oxford (2000), novels are defined descriptively: printed works referred to as novels by contemporary readers. As *ibid.* note, this definitional strategy becomes particularly viable after 1770 as the characteristics of works referred to as novels stabilize. There is, however, a concern about the tendency of this definitional strategy to exclude literary works plausibly characterized as novels. But different definitions of the novel tend to agree on particulars in most cases. And the cases where disagreement is likely to occur tends to be easy to predict. For example, a definition of the novel which tends to include novel-like juvenile fiction and novel-like religious fiction – both excluded by the definition in use in RFGS – would be likely to agree on particulars 90% of the time (Troy Bassett, personal communication, November 9, 2015).

A. Riddell (✉)
Indiana University Bloomington, Bloomington, USA
E-Mail: riddella@indiana.edu

M. Betancourt
Symplectomorphic, LLC, New York, USA

Antonia Forster,² Peter Garside, Rainer Schöwerling,³ Peter Garside et al.⁴ (hereafter ‘RFGS’) have already provided an exhaustive bibliography covering the years 1770 to 1836. The approach used to estimate novelistic production here is applicable to other geographic and linguistic contexts. Moreover, year-to-year changes in the rate of new novel publication in one region are likely to be informative about how quickly (or slowly) novelistic production is changing in neighboring regions. Yearly estimates of the rate of novelistic production are essential in a variety of projects associated with literary and publishing history. For example, these estimates facilitate bibliographic work as they permit literary historians to assess the completeness of existing bibliographies for a given year.

This contribution is usefully seen as participating in a broader project: a data-intensive, sociologically-inclined literary history seeking to address long-standing questions in the history of the novel using digital surrogates of surviving literary works.⁵ Although sociology of literature in the 1950s and 1960s foundered due to a lack of information about writers, works, and related documents—or representative samples thereof—contemporary library-scale digitization, inexpensive computational resources, and widespread sharing of machine-readable datasets permit resuming a variety of research projects on more secure footing.

Before presenting and discussing the new estimates of yearly novelistic production in the British Isles, this paper mentions existing research which stands to benefit from an exhaustive bibliography of the novel in the 19th century.

²James Raven/Antonia Forster, *The English Novel, 1770–1829: A Bibliographical Survey of Prose Fiction Published in the British Isles*, Vol. 1.: 1770–1799, eds. Peter Garside/James Raven/Rainer Schöwerling, Oxford (2000).

³Peter Garside/Rainer Schöwerling, *The English Novel, 1770–1829: A Bibliographical Survey of Prose Fiction Published in the British Isles*, Vol. 2: 1880–1829, eds. Peter Garside, James Raven, Rainer Schöwerling, Oxford (2000).

⁴Peter Garside/Anthony Mandal/Verena Ebbes et al., *The English Novel, 1830–36: A Bibliographic Survey of Fiction Published in the British Isles*, Cardiff/Paderborn (2003).

⁵The practical barriers to this genre of research have been slowly disappearing. Before the advent of library-scale digitization in 2002, accessing a representative sample of 19th century novels would have been prohibitively expensive and time-consuming. (Google Inc.’s book digitization project was launched internally at Google in 2002 *History of Google Books*, <https://books.google.com/intl/com/googlebooks/history.html> [retrieved on April 25, 2018]) With print runs often in the very low thousands, surviving works exist, often in small numbers and often in special collections, in libraries across North America and Europe. In this setting, the only way to access a book was to be in physical proximity to the work. This situation has improved remarkably in the last decade. The Internet Archive, the most significant institution providing unrestricted access to scans of books from North American libraries, has digitized over 2.9 million books from North American libraries alone and continues to add thousands of scans a month. On August 17, 2016 the *Internet Archive* had 2,375,270 scans in its North American libraries collection and 531,641 scans in its Canadian Libraries collection (<https://archive.org> [retrieved on April 25, 2018]).

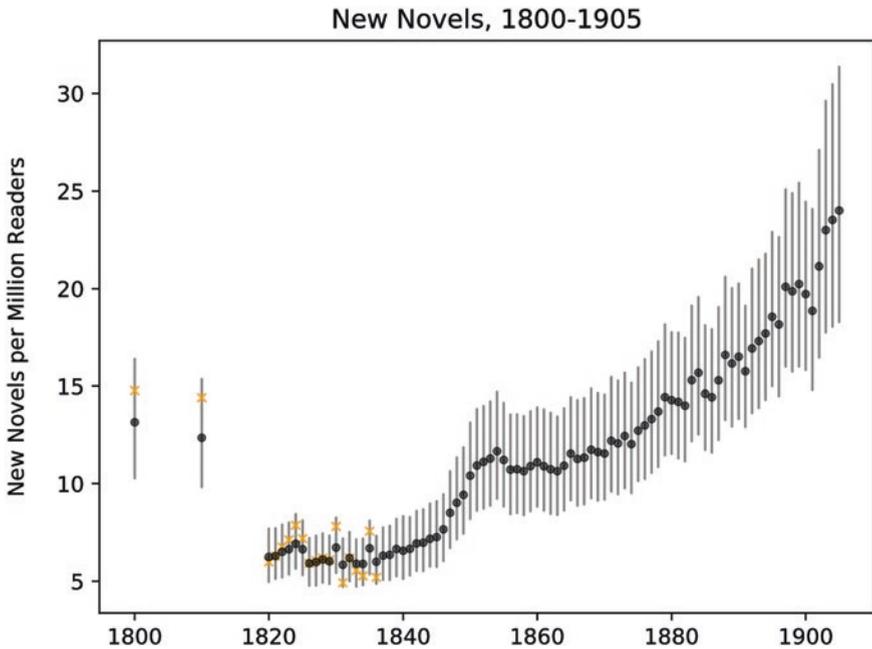


Fig. 1 New Novels Per Million Readers published in the British Isles⁶

2 Related Work

Projects likely to benefit from the availability of an exhaustive bibliography of published novels tend to be affiliated with the sociology of literature or sociologically-inclined literary history. Sociologically-inclined literary history has long been associated with inclusionist demands as well as the use of new methods. To the extent that previous research agendas linked to sociology of literature

⁶Cross marks indicate known data from RFGS. Points are median estimates and error bars indicate 80% credible intervals. The model used for estimating intervals is described in the Methodological Appendix. UK Population figures are from Angus Maddison (*Statistics on World Population, GDP and Per Capita GDP, 1–2008 AD*, n.p. 2009). Literacy figures are collected in Gregory Clark (*A Farewell to Alms: A Brief Economic History of the World*, Princeton 2007). Population figures prior to 1820 are only available for 1800 and 1810. Especially prior to 1820, population and literacy estimates are somewhat unreliable. Adjusting for population and literacy is, however, essential due to the pace and scale of change. Between 1800 and 1900 population likely increased by between 300 and 400%. Basic literacy in 1800 was likely around 61% for men and 42% for women. By 1900 literacy was approaching 100%.

failed or were abandoned due to the lack of reliable information about novelistic production, these agendas merit revisiting for reasons mentioned earlier—notably the relative abundance of book page images and book reviews as well as machine-readable bibliographic and biographical records. These agendas hold out the promise of studying literature at multiple scales and of enlarging the available vocabularies for discussing the history of literature.

A characteristic demand of sociologically-inclined literary history is the demand for an inclusive literary history. The most recent advocate for an inclusive history of the novel is Franco Moretti.⁷ Of the ca. 25,000 new novels published in the British Isles during the 19th century, Moretti estimates only about 200—less than 1%—figure in teaching and research.⁸ The demand for an inclusive approach to literary history and many research questions discussed under the heading of sociology of literature are present in earlier works as well, in particular Daniel Mornet's work from the 1910s and in Robert Escarpit's *Sociologie de la littérature* (1958).⁹ Mornet embraced an inclusive approach to the study of literary taste, explored patterns in reading broadly, and suggested attending to the potential influence of 'minor' authors. The research agendas of Escarpit, Mornet and sociologically-inclined literary history in general suffered due to the lack of accessible and trusted bibliographic data.¹⁰ The resources available to them did not include, needless to say, machine-readable bibliographic records of all surviving 19th-century books much less digital facsimiles of the pages of tens of thousands of surviving works.

⁷Franco Moretti: „The Slaughterhouse of Literature,” in: *MLQ: Modern Language Quarterly* 61 (2000), H. 1, 207–227; Franco Moretti: *Graphs, Maps, Trees: Abstract Models for Literary History*. London: Verso (2005).

⁸Franco Moretti, „The Slaughterhouse of Literature,” in: *MLQ: Modern Language Quarterly* 61/1 (2000), 207–227, here: 207.

⁹Robert Escarpit, *Sociologie de la littérature*, Paris (1958); Daniel Mornet, „Les Enseignements Des Bibliothèques Privées (1750–1780),” in: *Revue d'Histoire littéraire de la France* 17/3 (1910), 449–496.

¹⁰Robert Darnton, “Reading, Writing, and Publishing in Eighteenth-Century France: A Case Study in the Sociology of Literature,” in: *Daedalus* 100/1 (1971), 214–256 enumerates the shortcomings of Escarpit's statistics in *Sociologie de la littérature*: “To take 937 writers over 410 years [1490–1900] is to spread the sampling pretty thin – an average of 2.3 writers a year. Adding or subtracting a single man could shift the graph by 5% or more, [...]” *ibid.*, 217. By the standards of the time, Escarpit's coverage is impressive. Escarpit covers more than twice as many writers per year as Raymond Williams, who in *The Social History of English Writers* considers 350 writers born between 1740 and 1920 (so, less than 1 writer a year) Raymond Williams, *The Long Revolution*, Harmondsworth, Middlesex, 1965, 254. (Williams used the *Oxford Introduction to English Literature and the Dictionary of National Biography*.)

The testimony of John Sutherland¹¹ to the lack of available bibliographical and biographical details is informative. Sutherland mentions the “sheer unavailability of necessary empirical knowledge” as an obstruction to literary sociology, adding, “[O]ne of the things that makes literary sociology so easy to do at the moment is that we don’t know enough to make it difficult.”¹² These sentiments echo Eliot’s description of difficulties encountered in his work during the late 1990s.¹³

A second characteristic of sociologically-inclined literary history, tied to the demand for an inclusive literary history, is an openness to new methods. The demand for a diversification of methods—in particular, beyond close reading (“direct textual reading”)¹⁴—is familiar in Moretti’s work and one linked explicitly to the need to analyze the morphology of the hundreds of thousands of

Table 1 New Novels by Decade¹⁵

Decade	Novels (RFGS)	Novels (model)	NSTC (LOCED)	Novels (model) (% NSTC)
1790–1799	704			
1800–1809	778	593–916		
1810–1819	669	558–864	34.158	1.6–2.5
1820–1829	829	644–978	45.470	1.4–2.2
1830–1839	approx. 765.0	739–1127	52.395	1.4–2.2
1840–1839		1011–1598	62.760	1.6–2.5
1850–1859		1579–2532	83.080	1.9–3.0
1860–1869		1934–3101	83.080	2.3–3.7
1870–1879		2648–4226		
1880–1889		3840–5976		
1890–1899		5382–8330		
1900–1909		7724–13.306		
1910–1919		9374–19.815		

¹¹ John Sutherland, „Publishing History: A Hole at the Centre of Literary Sociology,” in: *Critical Inquiry* 14/3 (1988), 574–589.

¹² *Ibid.*, 588.

¹³ Simon Eliot, “Very Necessary but Not Quite Sufficient: A Personal View of Quantitative Analysis in Book History,” in: *Book History* 5/1 (2002), 283–293, DOI <https://doi.org/10.1353/bh.2002.0006>.

¹⁴ Franco Moretti: “Conjectures on World Literature,” in: *New Left Review* 1 (2000).

¹⁵ Sources: 1779–1836, RFGS; ‘Novels (model)’ shows the 80 percent credible interval predicted by the model; *Nineteenth-Century Short Title Catalog (NSTC)* is NSTC (LOCED) from Eliot (1997). The duplicate values in NSTC (LOCED) for the 1850s and 1860s are not typos; the number of titles associated with each decade are the same.

literary works.¹⁶ The openness to a mixture of methods in sociologically-inclined literary history is easy to distinguish from recent forms of interdisciplinarity in literary studies because, as observers have noted, the latter is sharply constrained: methods are allowed safe passage into literary studies only if they do not use numbers.¹⁷ Sociologically-inclined literary history, by contrast, entertains the borrowing of survey methods from the social sciences and a range of techniques from statistics, computational linguistics, complex systems, and biological systematics.¹⁸



Fig. 2 New novels published by largest four publishers¹⁹

¹⁶Moretti (Ann. 6), 208–209.

¹⁷James F. English, “Everywhere and Nowhere: The Sociology of Literature After ‘the Sociology of Literature’,” in: *New Literary History* 41/2 (2010), v–xxiii, DOI <https://doi.org/10.1353/nlh.2010.0005>, xii.

¹⁸Franco Moretti, *Graphs, Maps, Trees: Abstract Models for Literary History*, London (2005); Franco Moretti, „Network Theory, Plot Analysis,” in: *New Left Review* 68 (2011), 80–102; Janice A. Radway, *Reading the Romance: Women, Patriarchy, and Popular Literature*, Chapel Hill (1984).

¹⁹Calculations rely on manually identifying the publisher using the ‘publication’ field in a random sample of 360 titles published between 1810 and 1836 appearing in RFGS. Error bars (80% credible intervals) reflect uncertainty due to sampling. Yearly rate is calculated by dividing the number of novels published in each period by the number of years in the period.

3 The Rise of the Text Industry

Between 1840 and 1860 the rate at which new novels were published in the British Isles grew approximately twice as fast as the population of individuals able to read ('readers') (Fig. 1).²⁰ (The method used to estimate the rate of growth during this period are described in the Methodological Appendix.) Such a development would have required a doubling of numerous processes of interest to literary historians, not least a doubling of the number of publishable manuscripts. Existing narratives of the period do not observe this growth or, to the extent that they do, attribute growth to factors such as changing tastes among readers and clever marketing strategies by publishers. This section argues that factors such as declining unit costs and changes in the industrial organization of publishing deserve to play a



Fig. 3 Proportion of new novels published by largest four publishers²¹

²⁰The rate of publication of new novels grew at twice the rate of growth of the population and roughly twice the rate of the reading public. In 1820 approximately 6.0 new novels were published per million readers. By 1850, 10.5 new novels per million readers were appearing each year.

²¹Calculations rely on manually identifying the publisher using the 'publication' field in a random sample of 360 titles published between 1810 and 1836 appearing in RFGS. Error bars (80% credible intervals) reflect uncertainty due to sampling.

leading role in the narrative of the growth in novelistic production in the 19th century.

The pace of growth in new novels outstripped that of book publishing in general which itself was growing faster than the reading population between 1840 and 1860.²² Cataloging the yearly production of new titles is an obvious first step towards addressing a range of unanswered questions concerning the material processes by which literature is produced, distributed, marketed, and consumed. As Sutherland²³ notes, the extent of scholarly ignorance on these matters is near total. For example, nobody knows how many people pursued careers as novelists in the British Isles in the 19th century.²⁴ Narrow intervals containing the number of new titles published each year allow us to begin to answer these and related questions.²⁵

Between 1840 and 1860, the rate at which new titles appeared grew appreciably. After 1860 growth relative to the reading population is distinctly slower. It is something of a mystery, then, why the rate of new novel publication, adjusted for the reading population, would rise dramatically between 1840 and 1860. Such change involved a variety of other processes of interest: a doubling of manuscripts received, a doubling of paper consumed by publishing books, a doubling of the labor of compositors, and so forth. Given the fixed costs associated with writing or printing a novel, the increase in the number of new novels appearing in the literary market merits a thorough accounting. How was this extra demand (on the publisher's side) for manuscripts met? Assuming that the demand could not be met entirely by the existing population of novelists, where did the new novelists come from? Did their recruitment diversify the socio-economic or regional background of the existing pool of writers? (Table 1)

Two developments deserve to be part of any account of the increase in the rate at which new novels were being published. These developments have tended not to feature prominently in stories of the growth in new novels during the period or have tended to be equated with other factors such as changing tastes and clever advertising strategies. Both developments would have contributed to a decline in

²²The proportion of all editions which were new novels is increasing during this period. Figure 2.3 in Weedon (2003) shows the growth of titles in *Publishers' Circular* and NSTC relative to the number of people able to read Alexis Weedon, *Victorian Publishing: The Economics of Book Production for a Mass Market, 1836–1916*, Aldershot, UK, (2003), 50.

²³Sutherland (Ann. 10).

²⁴Ibid., 574–575; John Sutherland, *Victorian Fiction: Writers, Publishers, Readers*. Houndmills, Basingstoke, Hampshire (1995), 151–164.

²⁵Evidence that accounting for the material conditions of publication can yield useful insights is available in existing research. For example, Lee Erickson, *The Economy of Literary Form: English Literature and Theindustrialization of Publishing, 1800–1850*, Baltimore (1996), argues declining paper costs contributed to poetry's decline and John Sutherland, „Chips off the Block: Dickens's Serialising Imitators,“ in: n.e., *Dickens and Other Victorians*, London (1988), 97–119, DOI https://doi.org/10.1007/978-1-349-19503-9_7 (retrieved on April 25, 2018), argues market forces frustrated imitators of Dickensian serialization.

the unit cost associated with publishing a novel. This, in turn, would have led to a higher rate of new novel publication. After describing the two developments, we will discuss how declining unit costs would have likely led to an increased rate of new novel publication.

The first development is declining factor costs associated with the emergence of steam-powered printing and steam-powered paper production. It is clear that the cost of paper and the cost of printing declined tremendously over the 19th century, although precisely when and where the decline occurred needs to be determined.²⁶ Steam-power was introduced in the production of paper around 1807 and steam-powered printing was introduced around 1814.²⁷ By 1825 half of all paper in England was made by machine.²⁸ If paper and printing costs were not already declining by 1830, it seems likely there was at least the expectation of their doing so in the near future. This is consequential because paper and machining costs made up a considerable fraction of the cost of publishing a book.²⁹ estimates suggest that in 1850 paper and machining costs, taken together, exceeded the cost of compositing in print run of 1,000.³⁰

New and maturing forms of industrial organization are the second factor likely contributing to declining unit costs. Something which increasingly resembles the modern publishing industry emerges between 1840 and 1860.³¹ Associated developments likely contributed to declining unit costs via economies of scale, declining overhead, accounting improvements, and declining cost of capital.

In 1844 regulation was introduced that required companies to post audited balance sheets with the Registrar of Companies, a requirement which did not directly affect publishers, which were typically organized as partnerships, but gives a sense of broader changes in the business environment. Weedon³² considers 1844 a turning point, writing that “from this time [1844] on recognizable and systematic accounting systems began to take shape in many publishing houses.”³³ (Capital was indeed difficult to raise in the 1830s and grew easier over time.³⁴) One piece of evidence which would be consistent with an account focused on economies of scale would be evidence suggesting that larger publishers were steadily increasing the number of new novels they published. Since we lack comprehensive bibliographies for the period of interest there is no way to

²⁶Weedon (Ann. 18), 66.

²⁷Ibid., 64; James Raven, *The Business of Books: Booksellers and the English Book Trade, 1450–1850*, New Haven, CT, London (2007), 224.

²⁸Weedon (Ann. 18), 64.

²⁹Ibid.

³⁰Ibid., 87.

³¹Raven (Ann. 25), 328 f.

³²Weedon (Ann. 18).

³³Ibid., 62.

³⁴Ibid.

demonstrate this directly. We can, however, show that there is no evidence of any significant decline in concentration among publishers before 1840 (Fig. 2 and 3). We should also recall that measuring the output of a publisher using new novels understates production—and, with it, economies of scale—as many publishers of novels began to reissue previously published novels. And reissuing novels was an activity which grew in significance only after 1830. For example, Richard Bentley, who is among the top four publishers for 1835–1836, published 126 reissued works in the *Standard Novels* series between 1831 and 1855, 19 of which were published (with Henry Colburn) before 1833.³⁵

One well-known anecdote suggests the potential magnitude of efficiency gains between 1840 and 1860. Dickens' *Dombey and Son* (1846) and *Our Mutual Friend* (1865) were printed in similar quantities (30,000 and 32,000 respectively). Adjusting for deflation, the unit cost of *Our Mutual Friend* was roughly half (58%) of that of *Dombey and Son*.³⁶ If this example is typical, there is little need to search for additional factors contributing to the acceleration observed between 1840 and 1860. Publishing a book simply becomes less expensive.

We should expect to see individual publishers publishing more books if per unit costs fall. Why this should be is not immediately obvious because nothing forces a publisher, given positive expected returns and the opportunity to publish more books at the same cost, to take advantage of the opportunity. In a competitive environment, however, a variety of pressures would favor the publisher which prints the greater number of titles. Since publishing more books would bring with it increased revenue, lower variability in revenue, and economies of scale, firms printing more books would tend to accrue more financial resources, something which tends to be a competitive advantage. Moreover, bringing a greater number of books to market is, by itself, a costly and advantageous signal of quality to readers and to potential writers. Although there's considerable uncertainty about when and by how much unit costs declined during the period, we should be more confident that publishers who elected to print more rather than fewer titles would be more likely to survive in a competitive environment. In light of this, we should anticipate observing a greater number of novels being published when unit costs decline.

Competing narratives which attribute the increase in the per-reader rate of new novel appearance between 1840 and 1860 to growing demand for novels in the population deserve to be viewed with scepticism. Such accounts suggest that the increase is due to one or both of the following: 1) an increase in the number of novel readers (in excess of what would be expected from a growing reading population) and 2) an intensification of reading among existing novel readers.

³⁵Roger P. Wallins, "Richard Bentley," in: Patricia Anderson, Jonathan Rose (eds.), *British Literary Publishing Houses, 1820–1880*, Detroit (1991, 39–52), here: 43 f.

³⁶Sutherland (Ann. 10), 19.

The precise mechanism proposed varies but can include, for example, publishers' 'stimulating demand' with clever advertisements or typography. It would also include the suggestion that publishers, individually or collectively, managed to appeal to "public taste" in a previously unknown manner which bolstered demand.³⁷ An expanding market for books is often simply assumed, although typically without clarifying how fast the expansion is occurring—in particular, if it is expanding faster than the reading population:³⁸ is not exceptional in mentioning "[t]he burgeoning mass market of the first half of the nineteenth century."³⁹

One straightforward accounting of the increase in new titles would look to an increase in the number of readers of new novels among the existing population of readers. Such an increase is easy to imagine if we are persuaded that not everyone who can read and afford access to novels does indeed spend time reading novels. The strongest reason to discount this narrative, however, is that time spent on reading is inelastic. Given a finite number of hours in each day, time spent on reading is constrained and, with it, the number of books one is able to read in a given year. That the population of readers might, collectively, increase time spent reading new novels by as much as ten percent between 1840 and 1860 is difficult to credit. And a lateral shift away from non-novel reading to novel reading is incompatible with the evidence: the rate at which all books (non-novels and novels) is also accelerating over much of the period 1840–1860.⁴⁰

Another accounting of the growth in new novels focuses on the improving economic fortunes of the population. Given the considerable expense of a new novel (ca. 21 s 6d) or an annual subscription to a circulating library (ca. 20s)—roughly a week's wages for an unskilled worker in 1850—it is certain that a large fraction of the reading population could not afford access to new novels before 1840. This narrative, however, is implausible because there is no evidence of rising wages until well after the acceleration in the rate of new novel publication. Contrary to expectations about a period firmly within the industrial revolution, typical incomes did not start to rise until after 1830 and then only at a

³⁷ Raven (Ann. 25), writing about the period leading up to 1840, discusses strategies used to "stimulate demand", including a growing "sophistication" in booksellers' advertising tactics *ibid.*, 269–270. Missing from this account is any way to assess the relative importance of these tactics in the growing sales of books. Were these tactics essential or marginally important? Absent booksellers' stoking of demand, would sales of novels have grown at half the rate that they did? It seems possible that in a competitive environment improvements in typography and product design might have had zero influence on overall demand. That is, if one firm begins advertising, competitors will face pressure to do so as well even if the advertising has a negligible (or negative) effect on the overall quantity of goods demanded. Here advertising only potentially influences *which* firms receive business – not how much business they receive.

³⁸ *Ibid.*

³⁹ *Ibid.*, 334.

⁴⁰ Weedon (Ann. 18), 50.

relatively slow pace: an average yearly rate of growth of 0.86% between 1830 and 1860.⁴¹

The theory that the population of existing novel readers intensified their reading of new novels after 1840 and that this new demand influenced the rate of publication of new novels is difficult to credit for reasons already mentioned. Time available for reading sharply constrains the scope for intensification of novel reading among those who were already avid novel readers in 1840. For those who did not read many novels, increasing the rate at which they read novels more than even ten percent would encounter considerable friction. New novels remained luxury goods and circulating libraries' subscriptions limited the number of novels which could be borrowed at any given time.

What about the export market for novels? Total book exports do increase during the 19th century but there does not appear to be any particular departure or development of interest prior to or during the 1830s and 1840s.⁴²

The rapid growth between 1840 and 1860 in the per-reader rate of new novel publication deserves greater attention than it has received. This period of growth is surprising because it occurs before mass literacy and the general growth in incomes associated with the industrial revolution. It also deserves attention from literary historians because the sustained rate of growth is high enough that it could not have occurred without widely-felt changes in a variety of processes at the heart of the literary market.

Novel writers, in particular, must have felt the consequences of a doubling in the rate of new novel publication. For example, a doubling of the rate of new novels published would require either an intensification of labor by existing writers, a broadening of the population of writers, or some combination of the two. Any of these would likely be associated with substantial and durable changes in the relationship between writers and publishers. For example, an intensification of work or a broadening of the population of writers might have favored certain kinds of writers or the production of certain kinds of novels (e.g., novels in certain genres or with certain formal features) if these novels were easier (in some sense) to produce.

4 Conclusion

The digitization of bibliographic data and surviving novels makes a variety of tasks involved in the study of literary history less time-consuming and less resource intensive. In many cases this development enables research that would otherwise be abandoned as impractical.

⁴¹Robert C. Allen., "Engels' Pause: Technical Change, Capital Accumulation, and Inequality in the British Industrial Revolution," in: *Explorations in Economic History* 46/4 (2009), 418–435, DOI <https://doi.org/10.1016/j.eeh.2009.04.004>.

⁴²Weedon (Ann. 18), 40 f.

With distance from the arrival of—and optimism surrounding—large-scale library digitization projects in the late 2000s, it is perhaps easier to reflect soberly on the necessary supports for doing data-intensive literary history at scale. Without some knowledge of how many novels were published each year—along with related information such as publisher concentration and the demography of novelists—is difficult to make use of digital facsimiles. This paper contributes a description of the growth of novelistic production during the long 19th century, and, in particular, new estimates of the number of novels published each year between 1837 and 1919. These estimates will facilitate current bibliographic work and support future research in sociologically-inclined literary history.

Acknowledgements We gratefully acknowledge Troy Bassett for his feedback, his work on *At the Circulating Library*, and for providing estimates of the number of new novels published in 1886, 1891, and 1894. We would like to thank Cosma Shalizi for valuable comments on an early version of this paper. We are also grateful to Franco Moretti for providing machine-readable novel publication data for years prior to 1800.

Methodological Appendix: Inferring the Rate of New Novel Publication, 1800–1919

The quantities which will be estimated are the underlying yearly rates at which new novels (i.e., previously unpublished novels) appeared between 1800 and 1919. The number of novels published between 1800 and 1836 (inclusive) are known and can be used to check the predictions of the model.

1. Data

New Titles 1800–1836: Garside/Schöwerling⁴³ and Garside et al.⁴⁴ (‘RFGS’).

Thanks to an exhaustive survey we know the number of new novels appearing between 1770 and 1836 (inclusive).⁴⁵ (Only the numbers associated with the years 1800–1836 are used by the model.) The bibliography by RFGS both provides the most important time series used by the model and anchors the effort more generally by providing a particular definition of the novel. The counts of new novel publications during these years will be denoted with the variable y_i where i indexes the year. For example, the first year, 1800, is associated with the index 1.

⁴³ Raven/Forster (Ann. 2); Garside/Schöwerling (Ann. 3).

⁴⁴ Garside/Mandal/Ebbes et al. (Ann. 4).

⁴⁵ *Ibid.*; Raven/Forster (Ann. 2); Garside/Schöwerling (Ann. 3).

A sequence of counts will be indicated by joining the starting index and ending index with a colon, e.g., $y_{1:120}$.

NSTC Titles 1800–1870 with London, Oxford, Cambridge, Edinburgh, and Dublin Imprint Locations (NSTC (LOCED))

Simon Eliot⁴⁶ extracted yearly totals of titles listed in the *Nineteenth-Century Short Title Catalog (NSTC)* whose publication information indicates the place of publication was either *London, Oxford, Cambridge, Edinburgh, or Dublin (LOCED)*. The difficulty encountered in using these figures is that for years after 1836 we do not know what percentage of NSTC (LOCED) titles are new novels. These percentages must be inferred. The NSTC (LOCED) counts are recorded in the sequence $n_{1:71}$, where the index indicates the year.

The NSTC assigns undated material to the nearest half-decade (to a year ending with a ‘0’ or a ‘5’).⁴⁷ Rather than modify the model to accommodate this idiosyncrasy, years which end in a ‘0’ or a ‘5’ are not used in the model.

Publishers’ Circular Yearly Totals for 1843 to 1919 (PC)

Publishers’ Circular yearly totals are counts of the number of new editions reported in *Publishers’ Circular* between 1843 and 1919. These totals are reported in Simon Eliot.⁴⁸ These counts are recorded in the sequence $p_{1:120}$, where the index indicates the year. No reliable counts are available before 1843, so the only the subsequence $p_{44:120}$ is used.

The difficulty with using the PC data is the same difficulty encountered with the NSTC (LOCED) titles. The percentage of PC editions which are new novels is unknown. Although the specific percentage is unknown, we do have a considerable amount of prior information. From one year to the next, for example, it seems a safe assumption that the percentage is unlikely to change dramatically. Even over a longer period, we can be confident that wild swings are highly unlikely. We do not anticipate that the percentage of all editions which are new novels would swing from 2% between 1820 and 1824 to 20% between 1845 and 1849.

Elicited Quartiles for 1886, 1891, and 1894

Following the procedure described in Paul H. Garthwaite, Joseph B. Kadane, Anthony O’Hagan,⁴⁹ we elicited from Troy Bassett information about the number of new novels published during three years 1886, 1891, and 1894. For each year, quartiles of the distribution matching Bassett’s beliefs about the total number of

⁴⁶ Simon Eliot, “‘Patterns and Trends’ and the ‘NSTC’: Some Initial Observations. Part One,” in: *Publishing History: Cambridge* 42 (1997), 79–104.

⁴⁷ *Ibid.*, 86.

⁴⁸ Simon Eliot, *Some Patterns and Trends in British Publishing, 1800–1919*, London (1994).

⁴⁹ Paul H. Garthwaite/Joseph B. Kadane/Anthony O’Hagan, “Statistical Methods for Eliciting Probability Distributions,” in: *Journal of the American Statistical Association* 100/470 (2005), 680–700, DOI <https://doi.org/10.2307/27590587>.

new novels published in that year were elicited. As editor of *At the Circulating Library: A Database of Victorian Fiction, 1837–1901*,⁵⁰ Bassett has a range of information valuable in making accurate estimates. For example, Bassett has precise information about lower bounds for the number of novels published during these years (i.e., the number of novels already recorded for these years in the database) and the rate at which new novels are currently being identified.

The elicited quartiles used are the following:

- 1886: 393.75, 481.25, 612.5
- 1891: 481.25, 525.0, 612.5
- 1894: 656.25, 787.5, 962.5

Because *At the Circulating Library* uses a broader definition of the novel than, the quartiles shown above reflect a discount of 12.5% from the original elicited quartiles. This number was chosen as an approximation of the discount reported by Bassett. Bassett states that between 10 and 15% of the novels recorded in *At the Circulating Library* would be excluded from RFGS.

Given the quartiles of a distribution, a probability distribution which has approximately the same quartiles can be found. In this case, Gamma distributions are used as the approximating distributions.

Athenaeum Reviews of Novels: 1860, 1865, ..., 1900: Ellen Miller Casey⁵¹

Ellen Miller Casey provides counts for the number of novels reviewed in *The Athenaeum* during nine years: 1860, 1865, 1870, 1875, 1880, 1885, 1890, 1895, 1900. (*The Athenaeum* was a London literary magazine published from 1828 to 1921.) This series shows a steep increase in the number of titles reviewed. Like every source of book reviews published at the time, these counts underestimate the total number of new novels published as not all new novels are reviewed. These nine counts are recorded in the sequence $a_{1:120}$ where the index indicates the year. Only values associated with the nine years are used.

2. Model

The model offers a narrative of novelistic production in terms of the yearly rate of new novel publication. This narrative can be broken down into two main pieces: an underlying trend and periodic deviations from the trend.

⁵⁰*At the Circulating Library: A Database of Victorian Fiction, 1837–1901*, <http://www.victorianresearch.org/atcl/> (retrieved on April 25 2018).

⁵¹Ellen Miller Casey, “Edging Women out?: Reviews of Women Novelists in the *Athenaeum*, 1860–1900,” in: *Victorian Studies* 39/2 (1996), 151–171.

The model assumes that the rate of new novel publication tends to grow at a fixed percentage each year during the period. The model also makes assumptions about deviations from this growth trend. Accommodating deviations is essential because there is no doubt that a variety of events affected the rate of (new) novel publications during the period—e.g., a cholera epidemic in the 1840s, a recession in the 1850s, and World War I. In order to make accurate yearly predictions of the number of novels, the model must be flexible enough to accommodate significant deviations from trend.

The growth trend and deviations are modeled using a Gaussian process.⁵² Instead of having year-specific deviations be independently distributed (as in conventional linear models), deviations are assumed to be correlated with each other in a manner specified by a Gaussian process. The Gaussian process is characterized by a mean function m and a covariance function k (sometimes called a *kernel*). The covariance function used in the model is the squared exponential covariance function. An important parameter in the covariance function is the characteristic length-scale l , which governs to what extent deviations in nearby years are correlated with each other.

In symbols, the yearly rates of new novel publication are modeled on a log-scale as follows:

$$\begin{aligned} \lambda_{1:120} &\sim \text{MultivariateNormal}(\alpha_\lambda + \beta_\lambda t_{1:120}, K(t_{1:120})) \\ k(x_i, x_j) & \end{aligned}$$

K is an $n \times n$ covariance matrix whose elements are specified by the covariance function k . An element (i, j) of K corresponds to $k(x_i, x_j)$. $t_{1:120}$ is a column vector of integers associated with the years (1 for 1800, 2 for 1801, ..., 120 for 1919). The rate of new novel publication in year i is $\exp(\lambda_i)$. Generic weakly informative priors are placed on all the parameters not explicitly mentioned above. The characteristic length-scale l_λ is modeled with a Gamma distribution placing 90% of its mass on values between 1 and 10. This prior distribution expresses the belief that deviations will tend to persist for between 1 and 10 years. Such a prior is, for example, consistent with the belief that a financial crisis might affect the rate of publication in the short term but would cease to influence publication rates in years which are more than ten years distant from the crisis.

The yearly rates of new novel publication $\exp(\lambda_1), \exp(\lambda_2), \dots, \exp(\lambda_{120})$ are connected to observations of the number of new novels published with a Negative Binomial sampling distribution. The total number of new novels published is known for years between 1800 and 1836 (inclusive). In symbols, this sampling model reads:

$$y_i \sim \text{NegativeBinomial}_2(\exp(\lambda_i), \phi_y) \quad i \in \{1, \dots, 37\}$$

⁵²Andrew Gelman/John B Carlin/Hal S. Stern et al., *Bayesian Data Analysis*, Boca Raton (³2013), 501–516.

The negative binomial distribution has a variety of parameterizations. *NegativeBinomial*₂ is parameterized by a mean (i.e., location) parameter and a parameter controlling dispersion.⁵³

A second Gaussian process models the yearly proportion of PC titles which are new novels. In order to use the Gaussian process, these proportions must be expressed on the log odds scale. (The log odds is the logarithm of the odds, $\log\left(\frac{p}{1-p}\right)$, where p is a proportion between 0 and 1.) In contrast to changes in the yearly rate of new novel publication, the proportion of PC titles which are new novels is expected to change slowly. Whereas an economic depression might affect the rate of new novel publication over a period of several years, it would be far less likely to affect the proportion of PC titles which are novels. In symbols, the yearly proportions are modeled as follows:

$$v_{1:120} \sim \text{MultivariateNormal}(\alpha_v + \beta_v t_{1:120}, K(t_{1:120})) \\ k(x_i, x_j)$$

The proportion of PC titles which are novels in the year with index i is $\text{logit}^{-1}(v_i)$. ($\text{logit}^{-1}(x)$, the inverse logistic function, inverts the transformation of a proportion to the log odds scale.) The characteristic length-scale l_v is modeled with a Gamma distribution placing 90% of its mass on values between 8 and 36. This prior distribution expresses the belief that deviations will tend to persist for between 8 and 36 years.

In symbols, the sampling model for the observed yearly counts of PC titles is the following:

$$p_i \sim \text{NegativeBinomial}_2\left(\frac{\exp(\lambda_i)}{\text{logit}^{-1}(v_i)}, \phi_p\right), i \in \{44, \dots, 120\}$$

Similarly, the sampling model for the observed yearly counts of NSTC (LOCED) titles is as follows:

$$n_i \sim \text{NegativeBinomial}_2\left(\frac{\pi_n \exp(\lambda_i)}{\text{logit}^{-1}(v_i)}, \phi_n\right), i \in \{x \vee x \in \{1, 2, \dots, 71\} \wedge (x - 1) \bmod 5 \neq 0\}$$

The NSTC (LOCED) counts are always higher than the PC title counts. We make the assumption that the yearly rate for NSTC (LOCED) counts is equal to the yearly rate for PC title counts multiplied by a constant factor, π_n .

The distributions elicited from Bassett are incorporated as follows:

$$\lambda_{87} \sim \text{Gamma}(278.36, 46.3) \\ \lambda_{92} \sim \text{Gamma}(1530.73, 245.4) \\ \lambda_{95} \sim \text{Gamma}(667.9, 101.8)$$

⁵³If Y is distributed according to a *NegativeBinomial*₂(μ, ϕ) distribution then $E(Y) = \mu$ and $\text{Var}(Y) = \mu + \frac{\mu^2}{\phi}$.

These distributions approximate the elicited quartiles described above. Note that λ_i is the logarithm of the yearly rate of publication. The approximating distributions therefore also use the log scale.

Finally, the counts of new novels reviewed in *The Antheneum* are incorporated as follows:

$$\begin{aligned} a_i &\sim \text{NegativeBinomial}_2(\pi_a \exp(\lambda_i), \phi_n), i \in \{61, 66, \dots, 101\} \\ \pi_a &\sim \text{Gamma}(15.5, 32) \end{aligned}$$

Because *The Antheneum* does not review all new novels published, the rate of new novel publication is discounted by a factor π_a . An informative Gamma prior placing 90% of mass on a value between 30 and 70% is used as it seems very likely that *The Antheneum* reviews a fraction of new novels each year lying in this range.

3. Inference

Posterior inference is performed using Stan.⁵⁴

References⁵⁵

- Allen, Robert C., “Engels’ Pause: Technical Change, Capital Accumulation, and Inequality in the British Industrial Revolution,” in: *Explorations in Economic History* 46/4 (2009), 418–435, DOI <https://doi.org/10.1016/j.eeh.2009.04.004>.
- Miller Casey, Ellen, “Edging Women out?: Reviews of Women Novelists in the *Athenaeum*, 1860–1900,” in: *Victorian Studies* 39/2 (1996), 151–171.
- Clark, Gregory, *A Farewell to Alms: A Brief Economic History of the World*, Princeton 2007.
- Darnton, Robert, “Reading, Writing, and Publishing in Eighteenth-Century France: A Case Study in the Sociology of Literature,” in: *Daedalus* 100/1 (1971), 214–256.
- Eliot, Simon, “‘Patterns and Trends’ and the ‘NSTC’: Some Initial Observations. Part One,” in: *Publishing History; Cambridge* 42 (1997), 79–104.
- Eliot, Simon, *Some Patterns and Trends in British Publishing, 1800–1919*, London 1994.
- Eliot, Simon, “Very Necessary but Not Quite Sufficient: A Personal View of Quantitative Analysis in Book History,” in: *Book History* 5/1 (2002), 283–293, DOI <https://doi.org/10.1353/bh.2002.0006>.
- English James F., “Everywhere and Nowhere: The Sociology of Literature After ‘the Sociology of Literature’,” in: *New Literary History* 41/2 (2010), v–xxiii, DOI <https://doi.org/10.1353/nlh.2010.0005>.
- Erickson Lee, *The Economy of Literary Form: English Literature and Theindustrialization of Publishing, 1800–1850*, Baltimore 1996.
- Escarpit Robert, *Sociologie de la littérature*, Paris 1958.

⁵⁴ Stan: A C++ Library for Probability and Sampling, Version 2.16. O. O. 2017, DOI <https://doi.org/10.5281/zenodo.814234>.

⁵⁵ All digital references were last viewed on April 25, 2018.

- Garside, Peter/Mandal, Anthony/Ebbes, Verena et al., *The English Novel, 1830–36: A Bibliographical Survey of Fiction Published in the British Isles*, Cardiff/Paderborn 2003.
- Garside, Peter/Raven, James/Schöwerling, Rainer, „General Introduction,” in: Peter Garside, James Raven, Rainer Schöwerling (eds.), *The English Novel, 1770–1829: A Bibliographical Survey of Prose Fiction Published in the British Isles*, Oxford 2000.
- Garside, Peter/Schöwerling Rainer, *The English Novel, 1770–1829: A Bibliographical Survey of Prose Fiction Published in the British Isles*, Vol. 2: 1880–1829, eds. Peter Garside, James Raven, Rainer Schöwerling, Oxford 2000.
- Garthwaite, Paul H./Kadane, Joseph B./O’Hagan, Anthony, “Statistical Methods for Eliciting Probability Distributions,” in: *Journal of the American Statistical Association* 100/470 (2005), 680–700, DOI <https://doi.org/10.2307/27590587>
- Gelman, Andrew/Carlin, John B./Stern, Hal S. et al., *Bayesian Data Analysis*, Boca Raton ³2013.
- Maddison, Angus, *Statistics on World Population, GDP and Per Capita GDP, 1–2008 AD*, n.p. 2009.
- Moretti, Franco, „Conjectures on World Literature,” in: *New Left Review* 1 (2000).
- Moretti, Franco, *Graphs, Maps, Trees: Abstract Models for Literary History*, London 2005.
- Moretti, Franco, „Network Theory, Plot Analysis,” in: *New Left Review* 68 (2011), 80–102.
- Moretti, Franco, „The Slaughterhouse of Literature,” in: *MLQ: Modern Language Quarterly* 61/1 (2000), 207–227.
- Mornet, Daniel, „Les Enseignements Des Bibliothèques Privées (1750-1780),” in: *Revue d’Histoire littéraire de la France* 17/3 (1910), 449–496.
- Radway, Janice A., *Reading the Romance: Women, Patriarchy, and Popular Literature*, Chapel Hill 1984.
- Raven, James/Forster, Antonia, *The English Novel, 1770–1829: A Bibliographical Survey of Prose Fiction Published in the British Isles*, Vol. 1.: 1770–1799, eds. Peter Garside/James Raven/Rainer Schöwerling, Oxford 2000.
- Raven, James, *The Business of Books: Booksellers and the English Book Trade, 1450–1850*, New Haven, CT, London 2007.
- Sutherland, John, „Chips off the Block: Dickens’s Serialising Imitators,” in: n.e., *Dickens and Other Victorians*, London 1988, 97–119, DOI https://doi.org/10.1007/978-1-349-19503-9_7.
- Sutherland, John, „Publishing History: A Hole at the Centre of Literary Sociology,” in: *Critical Inquiry* 14/3 (1988), 574–589.
- Sutherland, John, *Victorian Fiction: Writers, Publishers, Readers*. Houndmills, Basingstoke, Hampshire 1995.
- Wallins, Roger P., „Richard Bentley,” in: Patricia Anderson, Jonathan Rose (eds.), *British Literary Publishing Houses, 1820–1880*, Detroit 1991, 39–52.
- Weedon, Alexis, *Victorian Publishing: The Economics of Book Production for a Mass Market, 1836–1916*, Aldershot, UK, 2003.
- Williams Raymond, *The Long Revolution*, Harmondsworth, Middlesex, 1965.

Online Resources

- About Google Books, History of Google Books, <https://books.google.com/intl/com/googlebooks/history.html>.
- At the Circulating Library: A Database of Victorian Fiction, 1837–1901, <http://www.victorianresearch.org/atcl/>.
- Internet Archive, <https://archive.org>.
- Stan: A C++ Library for Probability and Sampling, Version 2.16. O. O. 2017, DOI <https://doi.org/10.5281/zenodo.814234>.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.





Diskussionsbericht Sektion IV. Schnittstellen

Constanze Baum

Eine Schnittstelle ist einerseits ein technisches Protokoll zur Nachrichtenübermittlung, andererseits bezeichnet sie im übertragenen Sinne die Übergangsstelle zwischen verschiedenen in sich geschlossenen Systemen. Durch definierte Rahmenbedingungen bzw. Standards ermöglicht sie den Austausch von Daten, vermittelt aber auch ein Verständnis zwischen unabhängigen, miteinander kommunizierenden Systemen. Digitale Schnittstellen bestehen auf den Ebenen ‚von Maschine zu Maschine‘ und ‚von Mensch zu Maschine‘. Man kann sie allgemeiner aber auch als Metapher der Vermittlung verstehen. Eine standardisierte Schnittstelle ist somit nicht nur erforderlich, wenn man fremde Daten integrieren oder eigene Daten anderen zur Verfügung stellen möchte, ohne genau wissen zu müssen, wie das jeweils andere System im Inneren beschaffen ist. Der Begriff der ‚Schnittstelle‘ ist auch geeignet, Eigenschaften zu beschreiben, die im Allgemeinen miteinander kommunizierenden Systemen zukommen. Damit ist das Thema ‚Schnittstelle‘ auf verschiedenen Ebenen prädestiniert, die funktionalen und technisch-konzeptionellen Rahmenbedingungen der *Digitalen Literaturwissenschaften* näher zu beschreiben.

Die Sektion IV hatte zum Ziel, unter diesem Thema, das u. a. auch in der Medientheorie intensiver diskutiert wird, vor allem die Frage der Standardisierung, der Interaktion und die Beziehung der *Digitalen Literaturwissenschaft* zur *Digitalen Sammlung*, dem Archiv und der Bibliothek als den Reservoiren, die die *Digitale Literaturwissenschaft* mit Daten und Dokumenten versorgen, zu adressieren. Dies war der Einsicht geschuldet, dass Digitalität mit neuen Suchmöglichkeiten einhergeht, mit Prozessen der Aggregation, der Interaktion oder auch Vernetzung, die ihrerseits aber nur dann wirksam werden, wenn

C. Baum (✉)
Humboldt-Universität zu Berlin, Berlin, Deutschland
E-Mail: Constanze.Baum@hu-berlin.de

sie in die Modellierungs- und Präsentationsverfahren der eigenen Forschung einfließen. Die Forderung nach Schnittstellenkonformität hat somit auf den verschiedenen Ebenen Auswirkungen auf den Forschungs- und Publikationsprozess und sollte in ihrer Tragweite und Virulenz diskutiert werden.

Diese Idee ließ sich nur teilweise umsetzen, was vermutlich darauf zurückzuführen ist, dass das Thema ‚Schnittstelle‘ in der Literaturwissenschaft – oder konkreter – die Probleme der Standardisierung oder Interaktion in der *Digitalen Literaturwissenschaft* in der täglichen Forschungsarbeit noch nicht in dem Maße im Vordergrund stehen, wie dies zu erwarten wäre, und sich bisher allenfalls als Ärgernis der Daten- und Dokumentbeschaffung sowie -nutzung aufdrängen. Insofern ist es auch überraschend, dass die *Digitale Literaturwissenschaft* gegenüber den Infrastruktureinrichtungen, der Bibliothek und dem Archiv, nicht nachdrücklicher die zentrale Bereitstellung und Aggregation digitaler Daten und Dokumente für ihre Forschungen einfordert. Die frei verfügbaren Korpora sind, folgt man den üblichen Klagen, entweder von mäßiger Qualität oder angesichts der analog vorhandenen Texte vom Umfang her immer zu schmal. Selbst *TextGrid*, *Gutenberg* und das *Deutsche Textarchiv*, die bei den deutschen Korpora immer wieder genannt werden, entsprechen weder quantitativ noch in Teilen qualitativ wissenschaftlichen Anforderungen. Andererseits sorgt die Literaturwissenschaft aber nur in begrenztem Umfang für die offene Bereitstellung ihrer Daten, geschweige denn publiziert sie selbst regelhaft offen unter freien Lizenzen in standardisierten Formaten, obwohl sie – wie auch die Beiträge von Martus und Jannidis gezeigt haben – sich selbst das Leben leichter machen würde, wenn die Zugänglichkeit der Daten und Dokumente, z. B. zur Reproduzierbarkeit der Ergebnisse, nicht durch rechtliche oder technische Hürden erschwert würden. All dies mag den Ausschlag gegeben haben, warum einerseits vergleichsweise wenige Beiträge des Symposiums sich selbst hier verortet, andererseits viele das Thema implizit oder explizit, gelegentlich auch unter anderen Überschriften, aufgegriffen haben. So kam in den vorangegangenen Sektionen z. B. die Problematik der Korpusbildung, deren Klassifikation und flexible fragenabhängige Zusammenstellung, die Vereinheitlichung der Annotation oder die standardisierte Struktur für Zeitschriften zur Sprache. Auch die Frage der Langzeitarchivierung, also des Reservoirs der Literaturwissenschaft, kam wiederholt auf, etwa bei den verschwundenen Geniusedaten oder bei der Archivierung von Netzressourcen wie lebendigen Blogs, ebenso bei der Sorge um die langfristige Pflege von Oberflächen *Digitaler Editionen*. Standardisierung schließt Fragen der Lemmatisierung oder des *POS Taggings* bzw. der Tokenisierung *Digitaler Editionen* ein, wie im Beitrag von Brüning oder eben allgemein in der immer wieder sich aufdrängenden Frage nach brauchbaren Korpora für das *Textmining* deutlich wurde. Dies ist vorauszuschicken, um die möglicherweise nicht immer ganz passgenaue Einordnung der in dieser Sektion versammelten Beiträge zu erklären.

In praktischer Hinsicht bestehen für die Literaturwissenschaft vielfältige Schnittstellenanforderungen – je nachdem, wo Aggregation oder Vernetzung notwendig sind. Dazu kann ein einheitliches TEI-Format gehören (z. B. *TEI simple*) oder aber die einheitliche Ansetzung von Entitäten (Personennamen, Orte,

Zeitbegriffe, Gegenstände). Schnittstellenanforderungen ergeben sich auch mit Blick auf den üblichen Kontext der Bereitstellung im Internet in der Beziehung der Dokumente zueinander (nach dem lateinischen Motto *liber librum aperit*), das auch in einigen Beiträgen zutage trat. Hier überraschte, dass die eigentlich naheliegenden Themen ‚Hypertextualität‘ und *Semantic Web*, die für viele Fragen der *Digitalen Literaturwissenschaft* Lösungen bereithalten, bislang immer noch eine untergeordnete Rolle spielen. Effiziente Schnittstellen bieten Kataloge mit Metadaten, die nicht nur die Datenübernahme in eigene Publikationen erlauben und bibliographische Daten zur Verfügung stellen, sondern darüber hinaus auch eigene bibliometrische Auswertungen ermöglichen. Archiv und Bibliothek sind Repositorien für Daten und Texte und stellen diese idealerweise der *Digitalen Literaturwissenschaft* in standardisierter und maschinenlesbarer Form zur Verfügung: Retrospektiv als *Digitale Edition* oder digitalisierter und re-kodierter Text der schriftlichen Überlieferung, prospektiv als *born-digital* oder elektronische Publikation. Sie bilden Infrastrukturen für geistes- und kulturwissenschaftliche Forschungsdaten, die das schriftliche kulturelle Erbe in eine maschinenlesbare, durchsuchbare und algorithmisch prozessierbare Form überführen, nachhaltig sichern und der Forschung zur Verfügung stellen, die idealerweise ihrerseits Daten und Texte digital aufbereitet und annotiert, um sie dann in eine archivische oder bibliothekarische Infrastruktur zurückzuspielen. Wesentliche Bedingungen dieses auch als *Research Data Lifecycle* beschriebenen Prozesses sind neben technischen und informationswissenschaftlichen Anforderungen der digitalen Infrastruktur Nachhaltigkeitskonzepte, Standardbildung, qualitativ hochwertige, zuverlässige Daten und Texte sowie deren zentraler Nachweis über Metadaten, freie Zugänglichkeit idealerweise über *Open-Access*-Lizenzen und eine im *DH*-Konzept verankerte kooperative und inter- wo nicht transdisziplinäre Wissenschaftskultur. In den Beiträgen zur Sektion, die der Frage nach der Schnittstelle als Relais zwischen Bibliothek und Forschung auf sehr unterschiedliche Weise nachgingen, spiegelten sich viele dieser intendierten Themen und Facetten.

Der Beitrag von Karina van Dalen-Oskam widmete sich als Teilergebnis einer größeren Studie zur Leseforschung – *The Riddle of Literary Quality* (2012–2019) – der Schnittstelle von Übersetzungserfahrungen im transnationalen Diskurs. Van Dalen-Oskam ging anhand eines Korpus zeitgenössischer, populärer Literatur der Frage nach der Wahrnehmung von Fiktionalität und Übersetzung nach. Eine zentrale Ausgangsüberlegung war, inwiefern Lesern solcher Texte bewusst sei, dass sie eine Übersetzung lesen. Dem lag die produktive Unterstellung zugrunde, dass sich eine Differenz zwischen Ausgangssprache und Übersetzung als Leseerlebnis herstellen lasse. Von der Zielsprache abweichende Sprachmuster oder eine spezifische Anlage der Geschichte könnten demnach Hinweise sein anzunehmen, dass eine Übersetzung vorliege. Dass Textphänomene der Übersetzung als Qualitätskriterien wahrgenommen werden, erschien damit als notwendige Voraussetzung der Studie. Ob mittels der durchgeführten Umfrage ein *National Hiatus* zwischen zwei Sprachfamilien nachweisbar sei, war eine der Fragen, die sich in der Berichterstattung stellten; eine weitere daran anknüpfende Überlegung bezog sich darauf, welche Vergleichserhebungen nötig seien, um valide Daten zu

erhalten. In ihrer Replik stellte van Dalen-Oskam heraus, dass ihr bei der Wahl der Herangehensweise an der Kombination von soziologischer Feldstudie und literarischer Analyse gelegen sei. In der Diskussion des Beitrags wurde einerseits deutlich, dass das Thema Anschlussfähigkeit zu digitalen Methoden besitzt, die jedoch noch stärker gesucht werden müssten. Andererseits wurden auch eine Reihe nicht-digital basierter Aspekte im Hinblick auf das Thema diskutiert. So wurde über Lesekompetenz und die sie beeinflussenden Faktoren gesprochen, beispielsweise in welcher Korrelation Bildungsgrad und Leseverhalten zueinander stehen. Es wurde darauf hingewiesen, dass der Grad an Lesekompetenz wichtig für die Einschätzung von Ergebnissen sei. Zudem wurde nach den verschiedenen sozialen Milieus gefragt, aus denen niederländische Leser stammen und ob sich demzufolge Milieuunterschiede in der Bewertung der Übersetzungen abzeichnen. Kritisch wurde zudem eingewandt, dass zu prüfen sei, ob der vorgeschlagene komparatistische Ansatz ausreiche, das Phänomen adäquat zu erfassen. Da die Umfrage insgesamt ergeben habe, dass 11 % der Befragten nicht lesen, kamen auch Zweifel hinsichtlich der Konsistenz der Studie auf. Es wurden daher verschiedene Vorschläge unterbreitet, die Ergebnisse der Studie durch Interviews oder größere Korpora valider zu gestalten.

Das Problem, was eine gute literarische Übersetzung auszeichne und ob und wie Leser diese wahrnehmen, warf für van Dalen-Oskam größere Fragekomplexe auf: So gelte es herauszufinden, welche Faktoren die Literarizität eines Textes bedingen und – damit verbunden –, ob Literarizität digital messbar, erklärbar oder sogar programmierbar sei. Daran schließen sich Überlegungen nach dem literarischen Wert von Texten an. Kann dieser über normative Verfahren festgelegt und mittels digitaler Methoden erfasst werden? Die Bestimmbarkeit von Literarizität, so ein formulierter Einwand, könne sich nicht nur an vorgegebenen Kategorien orientieren. Literaturwissenschaft frage nicht nur nach den Konventionen. Werde ein hoher literarischer Wert impliziert, ginge es vielmehr auch um die produktive Lücke, die ein Text füllt, um literarisch zu wirken.

Die Diskussion zeigte auf, dass sich mit der Herangehensweise der Studie viele Fragen verbinden, deren komplexes Zusammenspiel möglicherweise noch nicht hinreichend definiert worden sei. So ließ sich nur schwer bestimmen, zu welchem methodischen Feld diese Form der *Digital Literacy Studies* zu zählen sei. Bekräftigt wurde aber, dass literatursoziologische Aspekte, wie sie die Leseforschung anbietet, in den Kontext der *Digitalen Literaturwissenschaft* eingebunden werden sollten und eine wertvolle Ergänzung zu anderen methodischen Herangehensweisen darzustellen in der Lage seien. Kritisch eingewandt wurde dennoch, dass zwischen den weit gefassten methodischen Ansätzen der Studie noch argumentative Lücken beständen, die es zu überwinden gelte. So wurde angeregt, sich dem Diskursfeld mittels stilometrischer Methoden zu nähern. Satzlänge oder Wortreichtum seien mögliche Parameter, die an die Untersuchungsergebnisse angelegt werden könnten, um qualitative Aussagen darüber zu erhalten, warum Probanden einen Text als literarisch bestimmten.

Van Dalen-Oskam wollte die Fragestellungen des Projekts in einen weiten Kontext eingespant sehen: Leseforschung und damit einhergehend

literatursoziologische Aspekte seien interessante, aus ihrer Perspektive relevante Faktoren für mögliche Lösungsansätze. Um aber herauszufinden, was Literatur ausmache, sei es wichtig, dass kollaborativ verschiedenste methodische Ansätze miteinander verknüpft würden. So könnten stilometrische Verfahren zwar Unterschiede aufdecken, dies sei aber noch nicht die Erklärung für alles. Sie machte sich dafür stark, die Ergebnisse nicht durch normative Setzungen zu verengen, und plädierte für eine größtmögliche Offenheit in der Anlage der Erhebung. Vor einer Befragung ohne determinierende Parameter wurde jedoch gewarnt. Auch seien die Kontextualisierungen des Lesens möglicherweise zu indifferent. Es mache einen erheblichen Unterschied, ob man danach frage, was gelesen werden soll, was Lesen überhaupt als Leseakt bedeute und was konkret gelesen worden sei. Hinsichtlich der Erschließung wurde als Desiderat festgehalten, dass Titel von Übersetzungen nur schwer in bibliothekarischen Systemen auffindbar seien. Ob sich Literarizität durch die aufgezeigten Verfahren sinnföällig extrapolieren lasse, blieb ungeklärter Restbestand der Diskussion.

Die Erfassung von Brief-Korrespondenzen im Projekt *correspSearch* bildete den Ausgangspunkt von Stefan Dumonts Vorlage. Digitale Projekte wie diese böten neben der stetigen Inventarisierung vorhandener Ressourcen, die deren Sichtbarkeit und Auffindbarkeit fördere, neue Möglichkeiten, Korrespondenzen in Relationen, Absender-Adressaten-Netzwerken oder Subkollektionen zugänglich zu machen. Dabei zeichneten sich zwei Wege der Datenpräsentation ab: auf der einen Seite zentrale Datenbanken sowie deren Bedeutung für Standards und Normierungen, auf der anderen Seiten dezentrale Initiativen und offene Plattformen, die ebenfalls dazu beitragen, digital aufbereitete Bestände zusammenzuführen und Zuarbeiten zur weiteren digitalen Erschließung zu leisten. Um Briefmetadaten aus Briefverzeichnissen zu aggregieren und in einem Austauschformat verfügbar zu machen, würden Minimal-Standards benötigt, die im Fall von *correspSearch* auf einem reduzierten TEI basieren. Die Zusammenführung von Daten in einem Webservice, der technische Schnittstellen anbietet, führe in der Nachnutzung dazu, dass das Lesen eines Briefwechsels in unterschiedlichen Editionen möglich wird. Eine solche digitale Informationsinfrastruktur biete eine Mikro-Meta-Infrastruktur, die aber auch Grenzen hat, da sie nur edierte Briefe und nicht die archivalische Überlieferung einbeziehe. Daran anschließend wurde die Frage aufgeworfen, wie sich der Widerspruch zwischen der Referenz auf die konkrete Brief-Edition bei gleichzeitig fehlendem Verweis auf das Werk als Abstraktum lösen ließe. Unklar bleibe, was das Projekt formal eigentlich sei: Handelt es sich um einen Webservice, oder ist diese Art der Präsentation von Daten bereits als *Digitale Edition* einzustufen? In seiner Replik ging Stefan Dumont auf die Schnittstellenprobleme hinsichtlich der GND-Datensätze ein. Es fehle an Workflows und Ressourcen in Bibliotheken, um Fehler innerhalb der Normsätze dezentral zu korrigieren oder Ergänzungen einzupflegen. Dumont forderte die Entwicklung von entsprechenden Verfahren, um die wissenschaftliche Community besser und effizienter in diese Workflows einzubinden. Kollaboration könne die Qualität der Daten verbessern helfen, lautete der Tenor seiner Erwiderung.

In der Diskussion wurde mehrfach das Anliegen des Projekts und das Angebot für die Forschung gelobt. Einigkeit herrschte darüber, dass Erschließungsarbeiten von Korrespondenzen eine wichtige Grundlage weiterer literaturwissenschaftlicher Arbeiten darstellten. Die durch das Projekt auf digitaler Basis beförderte Zusammenführung von Material wurde daher von allen Gesprächsteilnehmenden begrüßt. Gleichzeitig wurden aber auch die Problemhorizonte offengelegt, die sich mit der Zusammenführung, der Erschließung und der Standardisierung der Metadaten ergeben. Der Gründung globaler Archive seien Grenzen der Machbarkeit gesetzt, denn Entscheidungen über die Auswahl und den Zuschnitt von Daten für die jeweilige digitale Instanz, die eine solche Zusammenführung präsentiert – sei es Service oder Edition –, seien die bestimmenden Faktoren eines jeden Projekts, die in Einklang mit darüber liegenden Instanzen stehen und entsprechende Schnittstellen bereithalten müssten. Konkret sei zu dokumentieren, was die Genese des Projekts bedingt, welche Entscheidungen hinsichtlich der Festlegung von Metadatenschemata oder anderer Kriterien getroffen wurden und welche Korpora unter welchen Voraussetzungen in das jeweilige Projekt integriert wurden. Die Erschließung müsse demzufolge mit ihrer (technischen) Anschlussfähigkeit einhergehen. Dumont stellte dar, dass *Kalliope* beispielsweise restriktiv hinsichtlich der Aufnahme von Datenbeständen sei, da hier keine Dokumente aus Privatbesitz zugelassen würden. Auch Dokumente, die teilweise nur im Druck überliefert seien, fänden keinen Eingang, da *Kalliope* nur Originale aus Nachlässen verzeichne. Für die wissenschaftliche Befragung der Überlieferungskette einer Korrespondenz könnten aber gerade auch solche Dokumente von Relevanz sein. Solche Erschließungslücken bzw. die spezifischen Bedingungen der Auswahl und Präsentation von Daten seien vielleicht auch im digitalen Raum nicht zu vermeiden, sollten aber in der jeweiligen Edition von Daten- oder Metadatenmaterial mittels textkritischer Verfahren dargelegt werden.

Weiter wurde angesprochen, dass der Datenbereitstellung von unterschiedlichen Projektträgern eine grundsätzliche Problematik inhärent sei, die besonders an der je verschiedenen Art der Zugänglichkeit und dem Recht auf wissenschaftliche Weiterverwertung deutlich ablesbar sei. Bedauert wurde, dass Rechtefragen der Präsentation der Volltexte entgegenständen. Dem könne nur durch langfristige Entwicklungen im Bereich von *Open Access*-Strategien durch freiere Lizenzen entgegengewirkt werden, da ansonsten die benötigte kritische Masse an Material nicht erreicht werden könne. Wenn nur Partikel Eingang in das Projekt fänden, stelle sich die Frage nach der wissenschaftlichen und empirisch-maschinellen Verwertbarkeit, da Ungleichgewichte drohten, wenn die Erschließung zu unvollständig sei und daraus beispielsweise ein verzerrtes Bild eines Korrespondenzen-Netzwerks erwachse. Für die Arbeitspraxis im Aufbau solcher digitalen Instanzen eröffne sich auch das Problem, inwieweit das Projekt selbst die Daten einsammeln und auch pflegen könne. Es wurde allgemein begrüßt, dass künftig auch Daten aus gedruckten Editionen erhoben werden sollen. Entsprechende Verfahren seien bereits angestoßen worden, befänden sich aber noch in der Entwicklung bzw. im Aufbau. Die Anbindung dieser Aufgabe an Akademien erscheine sinnvoll,

um große Datenmengen einzupflegen. Gleichzeitig sei zu klären, welche Rolle (Forschungs-)Bibliotheken hinsichtlich solcher Zuschnitte künftig spielen können.

Dumont betonte, dass das Format zur Einspeisung neuer Daten in *correspSearch* nicht automatisiert werden solle; aus datenkuratorischer Sicht sei eine Bewertung und Prüfung der eingehenden Bestände weiterhin notwendig. Einzuberechnen sei bei jeder neuen Ressource der oftmals hohe Zeitaufwand zur Identifikation von Personen und Orten. Angeregt durch Überlegungen zur Abstufungsmodellierung für die aufzunehmenden Metadaten, wodurch Fassung, Werk oder gedruckte Ausgabe identifizierbar wären, wurden Vorteile und Umsetzbarkeit solcher Verfahren diskutiert. So wurde vorgeschlagen, die für die Druckwelt entwickelten RDA-Standards für die Datenmodellierungen anzuwenden. Zugleich wurde die Komplexität betont, die sich mit der Datenerhebung verbinde und mit der man konfrontiert werde, wenn beispielsweise einem Brief zwei Absender zugewiesen werden müssten. Schlussfolgernd aus dieser Debatte herrschte Einigkeit darüber, dass diese Aufgabe einen größeren Projektantrag zur Erschließung der Daten nahelege. Die Extrahierung von Metadaten aus gedruckten Quellen sah Dumont als Herausforderung, die mittels OCR geleistet werden müsse. Es wurde auch eingeworfen, dass Datenbanken, z. B. *eXist*, helfen könnten, Berufsstände aus den XML-Dateien auszulesen. Eine Wortmeldung verwies auf die Chancen, die eine Internationalisierung des Projekts bieten könnte, beispielsweise durch die Integration der Datenbestände von *Early Modern Letters Online (EMLO)*. Auch ließen sich möglicherweise weitere Briefdatenbanken mittels entsprechender technischer Schnittstellen integrieren, genannt wurden in der Diskussion der Melanchthon-Briefwechsel und die Korrespondenzen der Weimarer Klassik. Subdatenbanken würden so in die nächsthöhere Instanz ‚geschaufelt‘.

Die Vorlage von Thomas Ernst griff Aspekte der ersten Sektion der Tagung auf. Ihm war an der Untersuchung von einem Text-Werk-Verhältnis gelegen, letztlich an der Frage nach einem möglichen neuen Werkbegriff, den die digitale Literatur durch ihre Prozessualität gleichsam herausfordere. Als soziales Medium verstanden, setze der digitale Raum Autor-Leser-Verhältnisse in ein neues Bezugssystem zueinander, dem es Kontur zu verleihen gelte. Der Werkbegriff sei laut der zentralen These von Ernst daher einer Revision zu unterziehen – ein Punkt, der von der folgenden Diskussion lebhaft aufgegriffen wurde. Unter Einbezug medientheoretischer und diskursanalytischer Überlegungen ging Ernst von zwei ‚Zuständen‘ aus: Er konstatierte einerseits einen statischen Werkbegriff, der im Sinne des Produkts eines Urhebers als Endpunkt des künstlerischen Schaffensprozesses angesehen werden könne, dem er andererseits einen dynamischen Werkbegriff gegenüberstellte. Die Öffnung des Werkprozesses werde in der Literaturwissenschaft seiner Ansicht nach innersystemisch gesehen, durch Paratexte und Intertexte, die den Text als dynamisches Konstrukt kennzeichneten. Dieses von der Literaturwissenschaft diskutierte Modell finde er in *Social Reading*, *Litblogs* und *Fan-Fiction* widergespiegelt und dort noch enger gefasst: Durch Angebote wie das Kommentieren rückten Lese- und Schreibakt im

digitalen Raum näher zusammen. Zu fragen sei demnach nach einer Veränderung von Autorschaftsfunktionen und -modellen, die digitale Textsorten wie Blogposts oder Tweets betreffen sowie die damit einhergehende Rolle von Interaktivität im Feld literarischer Kommunikation. Zur Disposition stehe demzufolge auch, ob die Relationen von Text und Rahmen im digitalen Raum noch verfügbar und stabilisierbar seien und welche Parameter diese Relationen bestimmen. Dies schloss die Frage ein, ob der Werkbegriff hinsichtlich der untersuchten Texte überhaupt aufrechterhalten oder ganz fallengelassen werden sollte. In seiner Replik machte Thomas Ernst noch einmal deutlich, dass ihm an Grundfragen der Literaturwissenschaft gelegen sei, die in letzter Konsequenz auf die Gründung einer ‚Netzliteraturwissenschaft‘ hinauslaufen könnten. Bezugnehmend auf konkrete Fälle stellte er fest, dass Leser und Leserinnen in sozialen Medien durch Kommentare als Produzenten und damit gleichsam Autoren in Erscheinung treten würden und durch das Kommentieren angestoßen so gesehen eine Leser-Leser-Produktivität zu verzeichnen sei und sich somit eine intensiviertere literarische Kommunikation in diesen Medienformaten abzeichne. Die Literaturwissenschaft habe für die kollaborativen Schreibnetzwerke allerdings noch keine angemessenen Beschreibungsmodi entwickelt. Neben Mode-Blog, Posting oder Kommentar sei *Code Poetry* als Zusammenspiel von Maschine und Leser ein weiteres Beispiel, für das sich die Frage nach der Beschreibung der Urheberschaft von Maschineninterventionen stelle. Sind Mitschreibende in der *Wikipedia* Modifikatoren? Oder können Beiträge auf der Plattform *Genius* – in Rekurs auf den Beitrag von Alexander Nebrig – als Gamifikationen gelten? Wie lässt sich die Einbindung unterschiedlicher Medienformate in Digitaldokumenten wie Twitter-Nachrichten fassen? Thomas Ernst sah in der Interaktivität und Prozesshaftigkeit, die diese heterogenen Phänomene anbieten, eine Auflösung des Werkbegriffs begründet und fragte danach, wie dies in den verschiedenen Bereichen diskutiert werde. Er sprach dabei unterschiedliche Problem- und Gemengelagen an, deren Begrifflichkeiten und Grenzziehungen im digitalen Raum seiner Ansicht nach zur Disposition gestellt würden, und forderte gezielt die Diskussion über eine Standortbestimmung heraus. Wie lassen sich *Digitale Edition* und digitale Version beispielsweise terminologisch und methodologisch präzise voneinander trennen? Thomas Ernst war an einer Debatte um den Werkbegriff interessiert, den er im Bereich der vernetzten digitalen Literatur als in Auflösung begriffen sah.

Eine Wortmeldung hob das Engagement hervor, sich aus literaturwissenschaftlicher Sicht mit der terminologischen Terraingewinnung zu befassen. Jedoch wurde die Ansicht vertreten, dass als Ausgangspunkt Phänomene und Bedingungen von Textproduktion besser geeignet seien, als die Theoretisierung beim Werkbegriff anzusetzen. Insofern wurde für „mehr Genette als Barthes“ plädiert. Tenor der gesamten Diskussion war jedoch auch, dass der digitale Medienwandel auf die Begriffsfindung zurückfalle und sich hier ein spannendes, theoretisch-methodisches Arbeitsfeld für die Literaturwissenschaft eröffne.

In der Diskussion um den von Ernst kritisch gesehenen Werkbegriff wurden sehr verschiedene Positionen markiert. So wurde u. a. die Historizität des Werkbegriffs thematisiert, wobei sich folgende Punkte als zentral herauskristallisierten:

Zum einen wurde gefragt, ob die herkömmlichen Begriffe terminologisch hinreichen würden, um digitale Textphänomene adäquat zu benennen; zum anderen, ob die theoretisch orientierte Hinterfragung des Werkbegriffs an dieser Stelle produktiv gemacht werden könne. Es wurden neue Begriffe eingefordert, die ‚Autorschaft‘ ersetzen könnten, man sah sich aber zugleich mit dem Dilemma von Überlappungen und Überschneidungen konfrontiert, die mit neuen und alten Begriffssetzungen einhergehen. Ein Wortbeitrag brachte den Aspekt der ‚Werkgenese‘ ins Spiel, da dieser im Zuschnitt besser zu den Phänomenen von Textproduktionen und -kommunikationen im sozialen digitalen Netz passe und stärker auf die ökonomischen Bereiche ziele, die die Dinge werkförmig und zugleich marktförmig machten. Im weiteren Schlagabtausch wurden unterschiedliche Positionen deutlich: So widersprach Ernst den Einwüfen, die den Werkbegriff für nicht geeignet hielten, die Debatte um die genannten digitalen Phänomene zu befördern, auch mit Rekurs auf entsprechende rechtswissenschaftliche Debatten. Zuspruch erhielt Ernst insofern, als auch – so eine Stimme – eine Wiederkehr des Werkbegriffs in aktuellen *DH*-Diskursen konstatiert wurde. Ein anderer Kommentar bestritt jedoch die Bedeutung des Begriffs für das Digitale. Ein längerer Diskussionsbeitrag führte aus, dass der Werkbegriff nicht inkonsistent werden dürfe, und insistierte auf verschiedene Kategorisierungsoptionen, wonach zwischen ‚Text‘ als syntaktischer, ‚Werk‘ als semantischer und ‚Ausgabe‘ als materieller Kategorie zu unterscheiden sei. Eine solche Systematisierung sei für diese Debatte wichtig. Gleichzeitig erfahre der Werkbegriff Wucherungen und Anlagerungen und damit auch eine neue Aufmerksamkeit. In der Verkopplung von Akteuren innerhalb des literarischen Kommunikationssystems liege demzufolge vielleicht eine neue Attraktivität begründet, die sich mit dem Werkbegriff verbinde. Zugleich wurde aber auch vor der Wirkmächtigkeit des Begriffs, z. B. im juristischen Diskurs, gewarnt. Barthes‘ kritischer Werkbegriff reiche hier möglicherweise nicht hin, die neuen Entwicklungen zu beschreiben. In diesem Sinne wurde vorgeschlagen, die inhärente Dynamik der sozialen Medien als Autor-Leser-Kommunikation zu verstehen, für die weniger das Werk zentral in Anspruch zu nehmen sei als vielmehr die Werkhaftigkeit im Sinne eines Umgangs mit den Dingen.

Dass die Auseinandersetzung mit dem Textumfang einen Forschungsgegenstand und zugleich ein Desiderat darstellt, bildete den Ausgangspunkt der Diskussion um die Vorlage von Frank Fischer und Robert Jäschke. Der Beitrag fragte danach, inwiefern digitale Methoden helfen könnten, den theoretischen Überlegungen empirisch belastbares Material zur Seite zu stellen, um Forschungsfragen in dieser Hinsicht neue Kontur zu verleihen. Dabei stellten sich schnell grundsätzliche, epistemologische Fragen nach der Natur des Textumfangs überhaupt, deren relationale Maßeinheit ‚Seitenanzahl‘ in der quantifizierenden Vergleichbarkeit nur ein Faktor sei. Schon die Korpuswahl, die sich nicht auf Volltexte sondern auf die Metadaten der *DNB* stützt, bringe hinsichtlich der Operationalisierbarkeit dieser Forschungsfrage größere Schwierigkeiten mit sich. Verschiedene Ausgabenformate von Erstausgabe zu Taschenbuchausgabe würden hier ebenso eine Rolle spielen – wie in der Diskussion festgestellt

wurde, enthalten Taschenbuchausgaben eine oft weitaus höhere Wortmenge pro Seite – wie die Beobachtung von Störfaktoren (z. B. Werbeseiten), die bei der Erfassung des Seitenumfanges Berücksichtigung finden müssten. Die Provokation, die dem Thema innewohnt, lag schon in den Ausgangsüberlegungen begründet, welcher Erkenntniswert für die Literaturwissenschaft sich mit der Frage nach dem Umfang verbinde und ob die Seitenanzahl das adäquate Maß zur Beurteilung sei. Eine weitere produktive Unterstellung war die Frage, ob es sinnvoll sei, Bücher mit mehr als 1000 Seiten zu veröffentlichen – durchaus auch durch einen entsprechenden Eingriff in den Satzspiegel –, weil diese dann statistisch gesehen kanonfähiger sein könnten.

In der Diskussion wurde hervorgehoben, dass sich die kategoriale Verengung auf den quantitativen Umfang von qualitativen Textmerkmalen ebenso wie von Fragen der Leseerfahrung oder Lesedauer entkopple. Es wurde die Frage aufgeworfen, warum als Grundlage nicht die Wortzählung gewählt worden sei, da die in der Vorlage erhobenen Metadaten zu viele Fehler aufwiesen. Als Grundlage könnten dann auch E-Books gewählt werden. Neben solchen Einwänden wurde das Projekt von verschiedenen anderen Seiten als ein gelungener, impulsgebender Beitrag angesehen, sich in literaturwissenschaftlicher Hinsicht mit dem Text und seinem Umfang als Forschungsfrage auseinanderzusetzen. Umfang als Textphänomen hänge mit diversen Faktoren und Phänomenen der Textproduktion und -verwertung zusammen, womit sich ästhetische, ökonomische und pragmatische Ansätze verbinden ließen, die nicht ohne Weiteres auseinanderzuidividieren seien, im Gesamt aber zur Bewertung der Funktion von Büchern beitragen. Bücher als Medium seien daher produktive Werkzeuge und ‚Quasi-Trittbretter‘ von Texten. Ein Text wie Jan Brandts *Gegen die Welt* (2011) zeige paradigmatisch auf, wie Buchdicke und Literaturkritik Hand in Hand gingen. Es wurde jedoch zu einer Präzisierung in der Nomenklatur geraten und vorgeschlagen, eine Unterscheidung zwischen Textlänge, basierend auf der vorhandenen Anzahl an Zeichen, und Ausgabenlänge, basierend auf der Anzahl der Seiten, vorzunehmen.

Frank Fischer führte aus, dass Extremtexte von über 1000 Seiten Phänomene wie Textverwitterungen bereithielten, die sich einer empirischen Zählung widersetzen. Umfang sei in diesem Fall auch eine poetologische Herausforderung an die Quantifizierbarkeit. In der Diskussion stellte sich auch die Frage, ob das vorhandene bzw. genutzte Datenmaterial als Informationsgeber ausreichend sei, um belastbare Ergebnisse zu erzielen. Es wurde angeregt, mithilfe einer anzulegenden Datenbank, mittels Teststatistiken oder Kollationen valide Daten zu erheben. Dem wurde entgegengehalten, dass die Operationalisierung zu komplex sei, um dies vorzunehmen. Versuchsberechnungen seien aber möglich und würden bei entsprechend vorliegenden Lizenzen auch erweitert.

In der weiteren Diskussion wurden die Dimensionen der Fragestellung für die Literaturwissenschaft generell, aber auch für mögliche digitale Herangehensweisen hervorgehoben. So verwies man auf die Rolle der Leserperspektive und die Option, Aufmerksamkeitsspannen als Quantifizierungsmöglichkeit in die Untersuchung einzubinden. Daran knüpften sich Überlegungen, die sich mit der Antizipation von Texten überhaupt beschäftigten, wie beispielsweise das selektive

Lesen und die Frage, inwiefern sich dies im Hinblick auf digital rezipierte Texte verändere. Wenn beispielsweise E-Book-Reader anzeigen, wieviel Content noch zu lesen ist, so sei dies ein Übertrag der Seitenzählung in ein neues und durchaus präsenters Maß, das Orientierung beim Lesen biete und den Blick auch auf Mengenwahrnehmung lenken könne. Fragen der unterschiedlichen visuellen Darstellung solcher Mengenangaben könnten daher auch in den Fokus der wissenschaftlichen Bewertung geraten. Die Diskussion legte weitere Desiderate offen, denen es sich im Hinblick auf das Phänomen des Umfangs nachzugehen lohnen könnte. Antworten seien beispielsweise auf die Frage zu suchen, welche Verlage die „dicksten Kinder“ produzierten und inwiefern Autor- und Verlagsreputationen hier als Steuermechanismen eingeplant werden müssten. Auf literaturhistorischer Ebene sprechen Text- oder Buchexperimente wie die der Gruppe *Oulipo* oder Lawrence Sternes *Tristram Shandy* dafür, dass das Verhältnis von Buch und Länge auch in der Poetik der Texte eine Rolle spiele und immer schon gespielt habe.

Im *Part-of-Speech Tagging (POS)* sei Länge in gewisser Weise eine Störvariable. So wurde auf Matthew L. Jockers verwiesen, wonach Länge zum Problem der *Digitalen Literaturwissenschaft* werden könne. Mit der Rückfrage, wie viel Länge Literatur brauche, wurde implizit auch auf die Frage nach der Relevanz von digitalen Verfahren wie beispielsweise dem *Topic Modeling* verwiesen.

Allen Riddell stellte in seiner Vorlage literaturhistorische Beobachtungen und informatisch komplexe Berechnungen zum Verhältnis von Buchproduktion und der Verbreitung über einen großen Zeitraum vor. Literaturhistorisch zeige die Studie ein Netzwerk von Kooperationen und belege vor allem die Anschlussfähigkeit von Literaturgeschichte an informatische Verfahren. Riddell fragte danach, was Gründe für Konjunkturen von Büchern seien. Durch die digitale Bestandsaufnahme werde eine Berechnung möglich und Phänomene der Buch- und Publikationsgeschichte könnten überhaupt erst sichtbar gemacht werden.

In der Diskussion wurde darauf hingewiesen, dass der sprunghafte Anstieg im Vertrieb in der Mitte des 19. Jahrhunderts in Verbindung mit anderen Kontexten, hier der industriellen Entwicklung, zusammenhängen könne. Die Gründung von Buchläden in Bahnhöfen im Zuge des Ausbaus der Eisenbahn könne so vielleicht ein historisches Argument liefern, den enormen Anstieg des Buchvertriebs zu erklären. Riddell wandte jedoch ein, dass die Berücksichtigung der ‚Railway Editions‘ den berechneten Effekt nicht hinreichend erklären könne. Die ökonomischen Bedingungen, die den weit abgesteckten Zeitraum bestimmen, seien als nicht stabil eingestuft worden. Interessant erschien es dennoch, der Frage nachzugehen, inwiefern Bücher im „goldenen Zeitalter der Buchproduktion“ verkauft worden seien und woher man Daten über Verkauf von Büchern im 19. Jahrhundert erhalten oder wie man diese rekonstruieren könne. In der Diskussion kam man schnell zu dem Schluss, dass Verfahren der *Digital Humanities* sinnfällig zeigten, wie Dinge für neue Wissenschaftsdisziplinen erschlossen werden könnten. Durch die Datenerhebungen würden Verbindungen von Geschichtswissenschaft und *Literary Studies* offengelegt, die es ermöglichten, transdisziplinär zu arbeiten.

Ein Diskussionsbeitrag betonte die Rolle von bibliographischen Ressourcen als Grundlage für die Kalkulation von verlorenen Büchern. Katalogsysteme seien in

dieser Hinsicht Stabilisatoren für die Überlieferungsgeschichte. Das Buch selbst könne verlorengehen, dies müsse aber nicht auch für den Katalog zutreffen. Er verzeichne als Seismograph von Exemplargeschichte auch die Verluste.

Abschließend wurde neben der grundsätzlichen Befürwortung einer Integration von Daten in andere Disziplinen ein Einwand vorgebracht, der mit den von Riddell gelieferten Zahlen und Berechnungen einherging und möglicherweise auch für andere Bereiche der *Digitalen Literaturwissenschaft* symptomatisch sei: „Können wir Berechnungen trauen, deren Operationen wir nicht mehr in Gänze verstehen oder nachvollziehen können?“ Kritisch eingeschätzt wurde demnach, welche Fragen sich also mit den gelieferten Daten und Informationen verbinden, und ob dies die Fragen seien, die es zu stellen gelte? Es wurde festgestellt, dass die Ergebnisse in dem von Riddell vorgestellten Bereich in eine Richtung zielten, die man als *Sociology of the Book* umschreiben könnte. Die Antwort auf die Frage, wie viele Exemplare eines Buches kursierten und wie es zirkulierte, müsse jedoch wieder stärker an den Text selbst rückgekoppelt werden. Auch Riddell räumte ein, dass geklärt werden müsse, welche Romane bzw. Bücher publiziert worden sind, um den Zahlen Kontur zu verleihen und dies in Korrelation mit der Anforderung zu bringen, welche Bedeutung ein Buch in einer spezifischen historischen Periode gehabt habe.

Der Schlussbeitrag von Thomas Stäcker regte zu einer grundsätzlicheren Diskussion über die Schnittstelle von Literaturwissenschaft und Bibliothek im digitalen Zeitalter an. Es wurde konstatiert, dass es sich hierbei um eine Konstellation handele, die durch den digitalen Wandel einem tiefgreifenden Transformationsprozess unterliege. Rollen müssten neu verteilt, möglicherweise überhaupt neu definiert werden. Die Doppelnatur der Bibliothek bedinge, dass ihr durch die zunehmende Durchdringung digitaler Ressourcen und Verfahren in allen Arbeitsprozessen Aufgaben zuwachsen, die sich stärker als zuvor mit der Forschung verknüpfen ließen. So könne die Bibliothek auf der einen Seite als möglichst effizienter Dienstleister der pragmatisch orientierten Nutzung auftreten, auf der anderen Seite sei sie nicht nur als Archiv, sondern vor allem als Laboratorium anzusehen, als humane Anstalt und schließlich Wissensnetzwerk. Wie aber definiert sich das Verhältnis von Bibliothek zu Literaturwissenschaft? Wo liegen Aufgaben, Heraus- und Überforderungen?

Die eingeschlagene Perspektivierung, die sich von der Bibliothek ausgehend zur Wissenschaft hin orientierte, ließ deutlich werden, dass hier anders gelagerte Parameter im Vordergrund stehen könnten: Das mit Digitalität verknüpfte Sammeln, Erschließen und die Nutzung würden mit der Anreicherung des Textes durch Daten, darunter der Einsatz von Markup-Verfahren, einhergehen. Text dürfe in dieser Perspektive nicht mehr in gleicher Weise als statisch angesehen werden. Vielmehr sei jeder bibliothekarische Akt bereits als Teil einer digitalen Forschungsfrage zu verstehen. Gleichzeitig wurde aber auch rückgefragt, welches Verständnis von ‚Digitalität‘ im Hinblick auf bibliothekarische Interessen eigentlich vorliege und inwiefern dies über Formen der Digitalisierung hinausreiche. Könnten Bibliothekssysteme im Rahmen der Katalogaufnahme beispielsweise auf Formen der maschinellen Autorschaft adäquat reagieren? Weitere Nachfragen

des Berichterstatters waren an die systemischen Aufgaben adressiert, die den Aufbau von *Digitalen Bibliotheken* im Zusammenhang von *Digitalen Editionen* und Tendenzen zu Zentralisierung und Dezentralisierung betreffen, auch im Hinblick auf verschiedene Sprachräume.

In seiner Replik betonte Thomas Stäcker zunächst, dass Ortlosigkeit sich auf Digitalität beziehe. Zur vorangegangenen Werk-Text-Debatte bemerkte er, dass die Existenzebene des Textes im digitalen Raum sei. Durch die Ablösung vom Trägermedium komme es zu einer Verschiebung, und die Zeichenrepräsentation werde daher aus seiner Sicht umso wichtiger. Stäcker verwies im Weiteren auf die Bedeutung von *Markup-Language* und XML. So sei beispielsweise Fettdruck eine Kulturleistung, die in Bedeutung übersetzt werden müsse. Text verweise demzufolge immer auch auf Struktur. Einem maschinellen Autor stehe er aufgeschlossen gegenüber, auch seien gleichermaßen künftig maschinelle Leser vorstellbar. Wer die Bibliothek befrage, so seine Conclusio, sei letztlich irrelevant, die Herausforderung liege vielmehr darin, dass die Bibliothek mit einer API technisch auf jede Frage antworten könne.

Hinsichtlich der Frage, welche Handlungspolitiken von zentralen Einrichtungen ausgehen sollten, betonte Stäcker den Kooperationsgedanken. Zugleich distanzierte er sich von dem Gedanken einer umfassenden Weltbibliothek im Digitalen. Anzustreben sei vielmehr eine Dezentralisierung, verteilte Zentren mit Spezialisierungsangeboten, um manövrierfähig zu bleiben. Künftige Aufgaben von Bibliotheken lägen nicht nur in der Bereitstellung, sondern auch in der Aufbereitung von Digitalisaten durch Volltextumwandlung (u. a. per OCR) und mit deskriptiven und strukturellen Metadaten. Hierfür sollten Verantwortlichkeiten klar benannt werden. Die Bibliothek stelle sich für Stäcker als Relaisstelle zwischen Allgemeinem und Speziellem dar, vor allem aber als zentrale Einrichtung im Prozess der *Datafication*. Durch die technisch-digitale Aufbereitung könne ein Maximum an Angeboten für die Nutzung, auch für forschersische Fragestellungen, erreicht werden. Allerdings bleibe es nicht aus, dass normative Setzungen in den Prozess der Bereitstellung und Aufbereitung eingreifen würden. So steuerten z. B. Vorschlagwortungen letztlich auch wieder Kanonisierungsprozesse.

Eine Wortmeldung plädierte für eine differenzierte Verwendung der Begriffe ‚Interface‘ und ‚Schnittstelle‘, wobei der Begriff ‚Interface‘ zu präferieren sei, der stärker das Aufeinander-Wirken herausstelle. Der Traum einer *Digitalen Bibliothek* werde in der Programmbibliothek verwirklicht, die hochwertig aufbereitete Texte bereitstelle. Zugleich gestand man ein, dass die *Digitale Bibliothek* ein *Moving Target* sei. Es ließe sich nicht klären, wo Forschung anfangen und wo es sich ‚nur‘ um Erfassung handele.

Stäcker griff dies auf und insistierte auf dem Aspekt der Interaktivität, der bislang zu wenig von der Forschung abgefragt worden sei. Die Bibliothek würde zu sehr reduziert auf einen Service von Angeboten, auf das Vor- und Aufbereiten. Sie sei aber Teil des Wissensdiskurses und auch die Recodierung des digitalen Erbes gehöre zu den Aufgaben der Bibliothek. Die Transformation ins Digitale beinhalte das Gestalten und Betreuen von aktuellem Wissen. Ob die Bibliothek künftig die Rolle von Herausgebern übernehmen werde und ob einer *Digitalen*

Bibliothek, wenn sie Forschungsdaten aufbereite, sogar eine Co-Autorschaft zuwachsen könne, waren weitere Aspekte der Diskussion, die auf die Frage nach dem veränderten Werkcharakter abzielten. Thomas Stäcker wies auf die medien-theoretischen Transformationen hin, die sich damit in materieller Hinsicht ver-bänden. Die Papier-Schrift-Kopplung und damit die Substratanbindung würden im digitalen Raum aufgegeben zugunsten eines anderen Verfahrens. Medien-theoretisch sei dies zunächst als Entzug zu bewerten. Dennoch sah auch Stäcker in diesem Entzug einen Gewinn für Neuformierungen: Werk- oder Textaus-gaben seien in bibliothekarischen Zusammenhängen dem funktionalen Zweck der Zusammenfassung unterstellt. Welcher Werkbegriff hinter einer solchen regulierten Erfassung stecke, sei damit nicht festgelegt. Er warnte zugleich davor, der Bibliothek alle Hoheitsrechte im Veröffentlichungsszenario zu überantworten. Vielmehr wurde festgestellt, dass es Forschung gebe, die inzwischen ohne Verlag auskomme. Die Bibliothek sei als Herausgeber demnach nur bedingt geeignet. Nicht beantwortet werden könne zudem die Frage, was die Aufgabe der Biblio-thek sein werde, wenn *Open Access* flächendeckend umgesetzt sei. Es fehle hier noch an Geschäftsmodellen, die im Brückenschlag von Bibliothek, Verlag und Autor adäquat die neuen Entwicklungen inkorporierten. Ergänzend wurde ein-gewandt, dass der Katalog des Bibliothekars – diskursgeschichtlich betrachtet – ein Instrument der Lenkung und Steuerung von Herrschaftswissen sei.

Was sich die Literaturwissenschaft von der Bibliothek wünsche, war ein weiterer Aspekt der Diskussion. Im Idealfall solle die Bibliothek in allen Über-setzungsfragen für die Forschenden aktiv sein und Mechanismen zur Verfügung stellen, um bibliothekarische Ressourcen auch für Forschungszwecke nutzbar zu machen (vgl. den Beitrag von Frank Fischer). So bliebe zu hoffen, dass Schnitt-stellen zu anderen Ressourcen wie *Wikidata* hergestellt werden. Im Hinblick auf das Zusammenwachsen bibliothekarischer und philologischer Interessen wurde angeregt, dass definiert und evaluiert werden müsse, wo die Prioritäten seitens der Wissenschaft liegen, um den Aufbau von Forschungsarbeiten sinn-voll zu unterstützen. Welchen Stellenwert habe beispielsweise die maschinelle Inhaltserschließung? Eine Rückkopplung an die Wissenschaft über den Wert solcher Automatisierungsprozesse wurde als überaus wichtig eingestuft. Auch schien es allgemeiner Konsens, dass die institutionellen Hürden für eine Partizipation niedriger gehängt werden müssten.

Stäcker schätzte den Prozess solcher Umsetzungen als steten Dialog ein, denn der Aufbau einer flächendeckenden Digitalisierung sei nicht generisch zu beantworten. Vielmehr, so wurde bemerkt, würden Forschungsaufträge vor allem über Förderformate und deren Gremien bestimmt. So sei durch die *DFG* beispiels-weise die Faksimiledigitalisierung der Verzeichnisse der im deutschen Sprachraum erschienen Drucke forciert worden (mittlerweile mit einer Deckungsquote von ca. 50 %). Die *DFG* als Vertreterin der Wissenschaft habe aus der Sicht Stäckers die Relevanz der Bereitstellung von Digitalisaten für die Forschung erkannt und konsequent betrieben.

Die Bibliothek als konkreter, aufsuchbarer Raum für Forschungsver-mittlung und ihre Auflösung im digitalen Raum standen im Mittelpunkt weiterer

Überlegungen. Hier wurden unterschiedliche Positionen laut. Eine Wortmeldung betonte, dass die *Digitale Bibliothek* ubiquitär zu denken sei. Es wurde nach Möglichkeiten der Operationalisierbarkeit gefragt, durch die Wissenschaftlerinnen und Wissenschaftler konkret Bedürfnisse einbringen könnten, um besser in Prozesse zur Ausgestaltung digitaler Zusammenhänge eingebunden zu werden. Zudem wurde ergänzt, dass Funktionen und Leistungen der Bibliothek vor dem digitalen Zeitalter übernommen beziehungsweise erhalten werden sollten. Auch im Digitalen seien Stätten für Forschungsfähigkeiten auszubilden. Die Bibliothek werde als Ort der Forschung gebraucht und von Akteuren auch als Lernort wahrgenommen. Der Ort ‚Bibliothek‘ gebe eine Idee davon, dass eine Übersicht möglich sei. Thomas Stäcker widersprach dem: In seinem Beitrag gehe es um die Medialität der Objekte, nicht um den Rahmen. Aus seiner bibliothekarischen Erfahrung heraus sei das Prinzip der forschersischen *Serendipity* durch die vage Nähe der Aufstellung in den Bibliotheken mittlerweile aufgekündigt, die Nutzer befänden sich nicht mehr vor den Regalen.

In seinem Schlussplädoyer stellte Stäcker noch einmal die Rolle der Bibliothek für die *Digitale Literaturwissenschaft* heraus. Dem Vorwurf, dass die Bibliothek durch Standardisierungen vorgebe, wie Editionen gemacht und die Freiheit der Edition durch solche Normierungen eingeschränkt würden, entgegnete er, dass Standardisierung der wissenschaftlichen Fachkultur auch im Analogen inhärent sei, um Wissenschaftskommunikation überhaupt möglich zu machen. Die Bibliothek sage nicht präskriptiv, wie Forschung zustande komme, es müsse aber ein Grundverständnis über Standards (z. B. zu Zitiermöglichkeiten im Netz) geben, um Forschungsergebnisse sichern und nachnutzen zu können. In diesem Sinne rief Stäcker abschließend dazu auf, sich über diese Parameter zu verständigen und neue Formate im Schulterschluss von Wissenschaft und Bibliothek zu definieren und auszudifferenzieren. Stäckers Apell als Bibliothekar richtete sich an die Disziplin: *Digitale Literaturwissenschaft* müsse mehr einfordern – vor allem von Bibliotheken –, damit bessere und umfangreichere Text- und Datensammlungen mit geeigneten Schnittstellen zur weiteren wissenschaftlichen Bearbeitung zur Verfügung gestellt würden.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

