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 revaluation 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, Belgien
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 revaluation 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.

---

[1] S. 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 viduity' (emphasis added); he winds back the tape, and then hears 'a-dying, after her long viduity' (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]

---

[2] Ruby Cohn, *A Beckett Canon*, Ann Arbor 2001.

[3] Samuel 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*.

[4] Cf. *BDMP*3, 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.

---

[5] Barbara 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.

---

[6] In 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 <del> 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 = "AF01|001">). 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

---

[9] James Joyce, *Finnegans Wake*, London 1939.

[10] Walton 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.

[11] James 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

---

[14] Theodor H. Nelson, *Literary Machines*, Sausalito 1981.

[15] Theodor 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

---

[16] Daniel Ferrer, "Quelques remarques sur le couple intertextualité-genèse," in: Paul Gifford/ Marion Schmid (eds.), *La création en acte*, Amsterdam 2007, 205–216, here: 206.

[17] Julia 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 *micro*analysis, involving careful transcription of barely legible manuscripts and the encoding of deletions, additions and marginalia, is the key to making future genetic *macro*analyses possible.

References[33]

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.

---

[30] Michael Riffaterre, "La trace de l'intertexte," in: *La Pensée* 215 (1980), 4.

[31] Jockers (Ann. 29), 4.

[32] Ibid., 171.

[33] All digital references were last viewed on May 20, 2018.

Ferrer, Daniel, "Quelques remarques sur le couple intertextualité-genèse," 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 degree*, 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.