Yearbook of Corpus Linguistics and Pragmatics

Jesús Romero-Trillo *Editor*

Yearbook of Corpus Linguistics and Pragmatics 2014

New Empirical and Theoretical Paradigms

Yearbook of Corpus Linguistics
and Pragmatics 2014

YEARBOOK OF CORPUS LINGUISTICS AND PRAGMATICS

2014

Editor-in-Chief:
Jesús Romero-Trillo
Universidad Autónoma de Madrid, Spain

Reviews Editor:
Dawn Knight, Newcastle University, Newcastle, UK

Advisory Editorial Board:

Karin Aijmer, University of Gothenburg, Sweden
Belén Díez-Bedmar, Universidad de Jaén, Spain
Ronald Geluykens, University of Oldenburg, Germany
Anna Gladkova, University of New England, Australia
Stefan Gries: University of California, Santa Barbara, USA
Leo Francis Hoye, University of Hong Kong, China
Jingyang Jiang, Zhejiang University, China
Anne O'Keefe, Mary Immaculate College, Limerick, Ireland
Silvia Riesco-Bernier, Universidad Autónoma de Madrid, Spain
Anne-Marie Simon-Vandenbergen, University of Ghent, Belgium
Anne Wichmann, University of Central Lancashire, UK

For further volumes:
http://www.springer.com/series/11559

Jesús Romero-Trillo
Editor

Yearbook of Corpus Linguistics and Pragmatics 2014

New Empirical and Theoretical Paradigms

Editor
Jesús Romero-Trillo
Departamento de Filología Inglesa
Facultad de Filosofía y Letras
Universidad Autónoma de Madrid
Madrid, Spain

ISSN 2213-6819 ISSN 2213-6827 (electronic)
ISBN 978-3-319-06006-4 ISBN 978-3-319-06007-1 (eBook)
DOI 10.1007/978-3-319-06007-1
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013940618

© Springer International Publishing Switzerland 2014
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

New Empirical and Theoretical Paradigms in Corpus Pragmatics, An Introduction 1
Jesús Romero-Trillo

Part I New Empirical Paradigms in Corpus Pragmatics

How Do Empirical Methods Interact with Theoretical Pragmatics? The Conceptual and Procedural Contents of the English Simple Past and Its Translation into French 7
Cristina Grisot and Jacques Moeschler

Subject Realization in Japanese Conversation by Native and Non-native Speakers: Exemplifying a New Paradigm for Learner Corpus Research 35
Stefan Th. Gries and Allison S. Adelman

***Jesus!* vs. *Christ!* in Australian English: Semantics, Secondary Interjections and Corpus Analysis** 55
Cliff Goddard

A Corpus-Based Analysis of Metaphorical Uses of the High Frequency Noun *Time*: Challenges to Conceptual Metaphor Theory 79
Shuangling Li

Part II Current Approaches to the Pragmatics of Culture and Society

Horace, Colors, and Pragmatics 99
Jacob L. Mey

Self-Conscious Emotions in Collectivistic and Individualistic Cultures: A Contrastive Linguistic Perspective 123
Barbara Lewandowska-Tomaszczyk and Paul A. Wilson

Translating Freedom Between Cultures and Ideologies: A Comparative Analysis of the Translation of Keywords in Galatians 149
Sarah Buchanan

How to Make People Feel Good When Wishing Hell: Golden Dawn and National Front Discourse, Emotions and Argumentation 179
Fabienne H. Baider and Maria Constantinou

Part III Advances in L2 Corpus-Based Pragmatics Research

'*We Went to the Restroom or Something*'. General Extenders *and Stuff* in the Speech of Dutch Learners of English 213
Lieven Buysse

Oral Production of Discourse Markers by Intermediate Learners of Spanish: A Corpus Perspective 239
Leonardo Campillos Llanos and Paula González Gómez

"Hope This Helps!" An Analysis of Expressive Speech Acts in Online Task-Oriented Interaction by University Students 261
Marta Carretero, Carmen Maíz-Arévalo, and M. Ángeles Martínez

Interaction and Codability: A Multi-layered Analytical Approach to Discourse Markers in Teacher's Spoken Discourse 291
Shanru Yang

Part IV Reviews

Review of Götz, S. (2013) *Fluency in Native and Nonnative English Speech*. Amsterdam: John Benjamins 317
Phoebe Lin

Review of Kretzschmar, W.A. Jr. *The Linguistics of Speech* (2009) Cambridge: Cambridge University Press 323
Paweł Szudarski

Review of Partington, A., Duguid, A. & Taylor, C. (2013) *Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies (CADS)*. Amsterdam: John Benjamins 329
Róisín Ní Mhocháin

Author Index 335

Subject Index 343

Contributors

Allison S. Adelman Department of Linguistics, University of California, Santa Barbara, Santa Barbara, CA, USA

Fabienne H. Baider Department of French Studies and Modern Languages, University of Cyprus, Nicosia, Cyprus

Sarah Buchanan Translation and Interpreting School of Modern Languages, Queen's University Belfast, Belfast, UK

Lieven Buysse Faculty of Arts, KU Leuven – University of Leuven, Brussels, Belgium

Leonardo Campillos Llanos Computational Linguistics Laboratory, Department of Linguistics and Modern Languages, Facultad de Filosofía y Letras, Universidad Autónoma de Madrid, Madrid, Spain

Marta Carretero Departamento de Filología Inglesa I, Facultad de Filología – Edificio A, Universidad Complutense – Ciudad Universitaria, Madrid, Spain

Maria Constantinou Department of French Studies and Modern Languages, University of Cyprus, Nicosia, Cyprus

Cliff Goddard School of Languages and Linguistics, Griffith University, Brisbane, QLD, Australia

Paula González Gómez Language Service, Universidad Autónoma de Madrid, Madrid, Spain

Stefan Th. Gries Department of Linguistics, University of California, Santa Barbara, Santa Barbara, CA, USA

Cristina Grisot Department of Linguistics, University of Geneva, Genève 4, Switzerland

Barbara Lewandowska-Tomaszczyk Institute of English Studies, University of Lodz, Lodz, Poland

Shuangling Li Department of English, University of Birmingham, Birmingham, UK

Phoebe Lin Department of English, Polytechnic University of Hong Kong, Hung Hom, Hong Kong

Carmen Maíz-Arévalo Departamento de Filología Inglesa I, Facultad de Filología – Edificio A, Universidad Complutense – Ciudad Universitaria, Madrid, Spain

M. Ángeles Martínez Departamento de Filología Inglesa I, Facultad de Filología – Edificio A, Universidad Complutense – Ciudad Universitaria, Madrid, Spain

Jacob L. Mey Institute of Language and Communication, University of Southern Denmark, Odense, Denmark

Róisín Ní Mhocháin TESOL, School of Education, Bath Spa University, Newton Park, Newton St Loe, Bath, UK

University of St Mark & St John, Plymouth, UK

Jacques Moeschler Department of Linguistics, University of Geneva, Genève 4, Switzerland

Jesús Romero-Trillo Departamento de Filología Inglesa, Facultad de Filosofíay Letras, Universidad Autónoma de Madrid, Madrid, Spain

Paweł Szudarski Department of English, Kazimierz Wielki University in Bydgoszcz, Chodkiewicza, Bydgoszcz, Poland

Paul A. Wilson Institute of English Studies, University of Lodz, Lodz, Poland

Shanru Yang School of Education, Communication and Language Sciences, Newcastle University, Newcastle, UK

New Empirical and Theoretical Paradigms in Corpus Pragmatics, An Introduction

Jesús Romero-Trillo

Abstract The second volume of the Yearbook of Corpus Linguistics and Pragmatics series, New Empirical and Theoretical Paradigms in Corpus Pragmatics, investigates the positive feedback between empirical and theoretical approaches to linguistics. The aim of the chapters in the volume is to propose new research paradigms to current studies in corpus linguistics and pragmatics.

The second volume of the Yearbook of Corpus Linguistics and Pragmatics series, New Empirical and Theoretical Paradigms in Corpus Pragmatics, investigates the indispensible positive feedback, though often transient in scholarly focus, between empirical and theoretical approaches to linguistics. The aim of the chapters in the volume is to propose new research paradigms to current studies in corpus linguistics and pragmatics, while trying to cross the threshold of traditional practices and venture into prospective paths of linguistic enquiry. The authors, with diverse backgrounds in pragmatics and corpus linguistics, offer novel theoretical and empirical models that can explain language better in itself and in its relation to reality.

To attain this aim, the chapters have been structured in four parts: (1) New empirical paradigms in corpus pragmatics; (2) Current approaches to the pragmatics of culture and society and (3) Advances in L2 corpus-based pragmatics research. The last section presents reviews of relevant volumes on different topics related to corpus linguistics and pragmatics.

The first part, New empirical paradigms in corpus pragmatics, opens with a chapter authored by Cristina Grisot and Jacques Moeschler entitled '*How Empirical*

J. Romero-Trillo (✉)
Departamento de Filología Inglesa, Facultad de Filosofía y Letras, Universidad Autónoma de Madrid, C/Francisco Tomás y Valiente 1, Madrid 28049, Spain
e-mail: jesus.romero@uam.es

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2, DOI 10.1007/978-3-319-06007-1_1, © Springer International Publishing Switzerland 2014

Methods Interact with Theoretical Pragmatics?'. Their study investigates an issue that has been present in theoretical pragmatics for several decades, i.e., the conceptual vs. procedural approach and its application to the description of verb tenses. With the aid of corpus-based methodology, the authors conclude that that the conceptual/procedural distinction should be investigated through empirical pragmatics in order to use real data, and thus incorporate the results to machine translation systems.

The chapter '*Subject Realization in Japanese Conversation by Native and Non-native Speakers: Exemplifying a New Paradigm for Learner Corpus Research*', written by Stefan Gries and Allison Adelman, presents the application of a new statistical method and its possibilities in Learner Corpus pragmatics research. It consists of a two-step regression procedure (MuPDAR) that determines the reason why some choices made by non-native speakers differ from those made by native speakers in a more comprehensive way than other traditional methodologies. They illustrate their theoretical proposal with the empirical analysis of Japanese native speakers' choices of grammatical, discursive and pragmatic realizations in comparison with non-native speakers of Japanese, and stress its advantages over other methods, specifically for error analysis purposes.

The next chapter, '*Jesus! vs. Christ! in Australian English: Semantics, Secondary Interjections and Corpus Analysis*', by Cliff Goddard, examines the different uses of such interjections and their nuances in pragmatic meaning in a corpus of Australian English. The methodological scope of the analysis incorporates the notion of speakers' ethnopragmatic awareness in context-bound elements, such as interjections, and how these elements can realize distinctive meanings that are amenable to semantic analysis through the Natural Semantic Metalanguage approach in liaison with corpus linguistics.

Shuangling Li ends the first section with the chapter '*A Corpus-Based Analysis of Metaphorical Uses of the High Frequency Noun Time: Challenges to Conceptual Metaphor Theory*'. Her chapter contributes to the ongoing debate on Conceptual Metaphor Theory and its validity with the application of corpus linguistics methodology. One of the most cutting-edge conclusions that the author reveals is that the collocational behaviour of certain lexical items and the phraseological uses of some linguistic expressions cannot be entirely explained, or at least have not been accurately described so far, in the traditional mappings of Conceptual Metaphor Theory.

The second part of the volume, Current approaches to the pragmatics of culture and society, departs with Jacob Mey's thought provoking chapter '*Horace, Colors and Pragmatics*'. In his contribution, and through the study of Horace's writings and of his apparent colour blindness, the author links the ethnological character of color terms in diverse societies, and by extension, how colors have different weight in the daily life of their inhabitants. Through a corpus-based analysis of Horace's texts (a truism, as all analyses of classical Latin are in essence corpus-bound) Mey adopts a socio-pragmatic approach to the analysis to aver the need for this perspective in the analysis of authors belonging to distant cultures or epochs.

Barbara Lewandowska-Tomaszczyk and Paul A. Wilson author the second chapter of this section entitled '*Self-Conscious Emotions in Collectivistic and Individualistic Cultures: A Contrastive Linguistic Perspective*'. In their contribution, the authors embark on the description of the relation between linguistic and culture-bound aspects of individualism and collectivism through the study of two emotions, shame and guilt, in English and Polish. Their study combines different theories with authentic corpus data, linking pragmatics with avant-garde theories on cultural identities. For this purpose, the authors use a questionnaire-based (GRID) methodology identifying cross-linguistic similarities and differences with respect to shame and guilt, validated with corpora, and then make interesting proposals for a more refined understanding of collectivistic and individualistic cultures.

The next chapter is Sarah Buchanan's '*Translating Freedom Between Cultures and Ideologies. A Comparative Analysis of the Translation of Keywords in Galatians*'. The author delves into the question of the conceptualization of values and beliefs in different cultures and societies, focusing on concepts such as freedom and slavery. Through a corpus-based analysis of Bible translations in German, French, Spanish and English, the author examines the translators' choices and the pragmatic implications of their decisions for readers of Paul's letter to the Galatians. This study, based on an intercultural corpus of sixteen translations, is a clear attempt to include corpus linguistics and pragmatics in the agenda of scholars working on translation and exegesis.

The last chapter of this section is authored by Fabienne Baider and Maria Constantinou and is entitled '*How to Make People Feel Good When Wishing Hell: Golden Dawn and National Front Discourse, Emotions and Argumentation*'. Their chapter, which has obvious social implications, analyzes the notions of resentment and hatred vis-à-vis the analysis of contempt and pride. By focusing on the pragmatic analysis of the metaphors and linguistic symbolisms used to embody these emotions in corpora, they illustrate how the use of these tropes leads to the social exclusion of some communities.

The third section of the volume, Advances in L2 corpus-based pragmatics research, starts with Lieven Buysse's contribution entitled '*We Went to the Restroom or Something. General Extenders and Stuff in the Speech of Dutch Learners of English*'. The study explores the use of general extenders, such as 'and stuff' and 'or something'. Through the comparison of two corpora, one of Dutch speakers of English and the other of native speakers of English. The study shows the general quantitative affinity of the phenomenon in both groups of speakers, but also the discrepancies in terms of fine-grained qualitative details. Some tentative explanations for the learners' choice of general extenders, i.e., L1 transfer, the intensity of exposure to certain forms in the target language, and learners' restricted repertoire of pragmatic devices, are suggested.

The next chapter, by Leonardo Campillos Llanos and Paula Gozalo Gómez, is entitled '*Oral Production of Discourse Markers by Intermediate Learners of Spanish: A Corpus Perspective*'. The authors investigate the oral production of discourse markers by learners of Spanish, and compare these results with those by native speakers. The overall description of the evolution of the markers in the

non-native speakers, together with a breakdown of the most used discourse markers in the corpus in native and non-native speakers, is used as a source of comparison and of some pedagogic guidelines for the teaching of Spanish.

Marta Carretero, Carmen Maíz-Arévalo and M. Ángeles Martínez contribute to the volume with a chapter entitled '*"Hope this Helps!" An Analysis of Expressive Speech Acts in Online Task-oriented Interaction by University Students*'. Their focus is on the presence of expressive speech acts in a corpus of e-forum history logs derived from online collaborative writing. Their hypothesis is rooted in the belief that in computer-mediated exchanges implicit disembodiment must lead to an outstanding role of expressive uses of language. The results of the analysis confirm the importance that expressive speech acts play in on-line communication.

The last chapter is authored by Shanru Yang and is entitled '*Interaction and Codability: A Multi-layered Analytical Approach to Discourse Markers in Teacher's Spoken Discourse*'. The author proposes a novel multi-layered analytical approach combining corpus linguistics, conversation analysis, and Second Language classroom modes analysis for the investigation of discourse markers in teachers' spoken language. The author exemplifies this approach through the detailed examination of macro and micro contextual uses of discourse markers in teacher-oriented classroom interaction in order to show the validity of this new methodology.

The volume ends with the reviews of three relevant volumes that epitomize significant and diversified updates in the study of corpus linguistics and pragmatics: Götz's (2013) '*Fluency in Native and Nonnative English Speech*'; '*The Linguistics of Speech*' (2009) by Kretzschmar & Partington; and Duguid & Taylor's (2013) '*Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies*'.

In sum, I believe that the contents of this volume with its new empirical and theoretical approaches to pragmatics and corpus linguistics, can be a very useful source of inspiration to understand the use of language in real contexts.

Part I
New Empirical Paradigms in Corpus Pragmatics

How Do Empirical Methods Interact with Theoretical Pragmatics? The Conceptual and Procedural Contents of the English Simple Past and Its Translation into French

Cristina Grisot and Jacques Moeschler

Abstract One major theoretical issue that has dominated the field of theoretical pragmatics for the last twenty years is the conceptual vs. procedural distinction and its application for verb tenses. In this chapter, we address this distinction from both theoretical and empirical perspectives following a multifaceted methodology: work on parallel corpora, contrastive analysis methodology and offline experimentation with natural language processing applications. We argue that the conceptual/procedural distinction should be investigated under the aegis of empirical pragmatics. In the case study, we bring evidence from offline experimentation for the procedural and conceptual contents of the English Simple Past and we use this information for improving the results of a machine translation system.

Keywords Empirical pragmatics • Corpus work • Linguistic experiments • Conceptual/procedural distinction • Natural language processing • Machine translation

1 Introduction[1]

In the last few years, linguists have become aware of the numerous advantages of the collaboration between theoretical and empirical pragmatics, which joined their forces in order to provide more and more insight into the use of language. In our view, empirical pragmatics investigates language use from both descriptive-theoretical and empirical perspectives. The empirical means considered in this

[1] The authors of this chapter are thankful to the reviewers for their helpful comments, which improved the quality of this chapter.

C. Grisot (✉) • J. Moeschler
Department of Linguistics, University of Geneva, 2 rue de Candolle 1211, Genève 4, Switzerland
e-mail: cristina.grisot@unige.ch; jacques.moeschler@unige.ch

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2, DOI 10.1007/978-3-319-06007-1_2, © Springer International Publishing Switzerland 2014

study are corpora and experimental methods. These methods are complementary and allow a better view on the linguistic phenomena of interest in this study, specifically the nature of the information encoded by verb tenses.

Theoretical pragmatics can be defined in a broad sense as the study of language in use, and in a narrow sense, as the study of how linguistic properties and contextual factors interact for utterances interpretation (Noveck and Sperber 2004). Two types of properties are involved in verbal communication: linguistic properties that are linked to the content of sentences (phonological, syntactic, semantic assigned by the grammar of each language) and non-linguistic properties that are linked to them being uttered in a given situation, at a given moment by a speaker. One question pragmatics wants to answer is the exact role of each type of property and their interaction. On the one hand, Grice (1975/1989) and neo-Gricean scholars (Gazdar 1979; Horn 1973, 1984, 1989, 1992, 2004, 2007; Levinson 1983, 2000) proposed an explanation based on conversation maxims and principles that guide conversation participants. On the other hand, relevance theorists (Sperber and Wilson 1986/1995; Blakemore 1987, 2002; Carston 2002; Moeschler 1989; Reboul 1992; Moeschler and Reboul 1994; Reboul and Moeschler 1995, 1996, 1998) speak about a unique expectation of relevance that hearers have while participating in an act of communication. According to relevance theorists, this expectation of relevance is sufficient for recovering the speaker's meaning.

Theoretical pragmatics (both neo-Griceans, relevance theorists as well as other pragmaticians) is thus concerned with phenomena related to the interpretation of utterances, including both explicit (in close relation to semantics) and implicit meaning. The main assumption is that propositional structures are systematically underdetermined and must be contextually enriched. Of great interest for the present study is the theoretical distinction between *conceptual* vs. *procedural* meaning, proposed by Blakemore (1987) within the framework of Relevance Theory (RT) (Sperber and Wilson 1986/1995). As Escandell-Vidal et al. (2011) argue, the conceptual/procedural distinction was first meant as a solution for the semantics/pragmatics division of labour and it has remained an important explanation for the contribution of linguistic meaning to utterance interpretation. A speaker is not expected to render more difficult than necessary his/her addressee's task in obtaining a relevant interpretation. Therefore, procedural meanings are instructions encoded by linguistic expressions that specify paths to follow during the interpretation process (manipulation of conceptual representations) in order to access the most relevant context. Wilson and Sperber (1993) attach cognitive foundations to the conceptual/procedural distinction and propose a distinguishing criterion: conceptual representations can be brought to consciousness while procedures cannot. We are particularly interested in this distinction because of its highly debated application for verb tenses (Smith 1990; Wilson and Sperber 1993; Moeschler et al. 1998; Moeschler 2000, 2002; de Saussure 2003, 2011; Amenós-Pons 2011; Moeschler et al. 1998, 2012; Grisot et al. 2012).

The two aims of this chapter are (1) to show that an investigation of the conceptual and procedural meanings of verb tenses should be done under the aegis of empirical pragmatics and (2) to argue for the benefits of combining two

empirical methods, corpus analysis and linguistic experiment. In our study, we combined data from parallel corpora that served as stimulus composition for offline experiments (linguistic judgement task). Parallel corpora revealed variation in translation possibilities of a verb tense from a source language (SL) to a target language (TL). Based on semantic and pragmatic theories we formulated hypotheses about the source of this variation and possible disambiguation criteria. Offline experiments allowed us to validate one of these criteria, as well as to propose new theoretic descriptions of the meaning and usages of verb tenses. We place this study under the cover of empirical pragmatics.

Empirical pragmatics draws on theoretical pragmatics and corpus linguistics, adopting experimental methods at the same time. Empirical pragmatics aims at having consistent data for supporting or challenging current pragmatic theories, as well as proposing new models for the interpretation of linguistic phenomena. Of course, theoretical pragmatics makes use of data consisting of built examples representing mainly the researchers' own intuitions. This type of data is criticisable mainly for its subjectivity and lack of replicability. For this reason robust (objective, quantifiable, replicable) data must be adopted, such as data from corpora (as argued for example by Barlow and Kemmer 2000; Boas 2003) and experiments (Tomasello 2000). Of the two types of experiments used in psycholinguistics, only offline experimentation can be adopted more easily by empirical pragmatics because of the lack of material required (no necessity of a laboratory with electroencephalography EEG material[2] or eye-trackers).

There is one branch of pragmatics that has integrated experimental methodologies for testing pragmatic theories: experimental pragmatics. While theoretical pragmatics is rooted in philosophy of language and in linguistics, experimental pragmatics, drawing on pragmatics, psycholinguistics and psychology of reasoning, has taken over and reinterpreted the psycholinguistic sophisticated experimental methods (Meibauer and Steinbach 2011). For instance, Katsos and Cummins (2010) emphasize the relation between pragmatic theory and psycholinguistic experimental design: linguists benefit from experimental data confirming the psychological validity of their observations and provide critical evidence for cases that go beyond the reach of intuitive reflection, and psychologists benefit from a wide range of phenomena to study and of multiple theories provided by semantics and pragmatics. Recent experimental pragmatics (such as papers from the volume edited by Noveck and Sperber in 2004) has focused on phenomena such as indirect speech acts, metaphors, implicature, presupposition and, more generally, speaker meaning.

Finally, we would like to argue that empirical pragmatics has built a bridge to the Natural Language Processing (NLP) domain thanks to the robust type of data used. The NLP domain needs models of language interpretation inspired from theoretical

[2] EEG is a procedure that measures electrical activity of the brain over time using electrodes placed on the scalp and it reflects thousands of simultaneously ongoing brain processes. Eye tracking is the process of measuring either the point of gaze or the motion of an eye relative to the head and it is used to investigate human thought processes.

pragmatics that can be adapted to machines. NLP also requires large amounts of data that allow quantitative analyses, statistical models and data for training parses and classifiers. Empirical pragmatics is able to provide NLP both linguistic models and empirical data.

This chapter is structured as follows: in Sect. 2, we introduce the role and type of data used in linguistics presented from a general point of view and in semantics and pragmatics, as well as their advantages and limits; in Sect. 3, we describe our case study by pointing out theoretical matters about verb tenses, our hypotheses, our empirical study on parallel corpora and offline experiments. We conclude our chapter in Sect. 4 by addressing the impact of the results of our experiments on theoretical matters about verb tenses and the importance of giving multiple sources of data for empirical pragmatics studies.

2 Type and Role of Data in Empirical Pragmatics

Nowadays, one can observe the increasing aspirations of linguists to use robust and objective findings in addition to intuitive and subjective acceptability judgements or built examples. McEnery and Wilson (2001) highlight that, broadly speaking, linguists have tended to favour the use of either introspective data (that is, language data constructed by linguists) or naturally occurring data (that is, examples of actual language usage). Nowadays, most linguists see these two types of data as complementary approaches, and not exclusive ones. Gibbs and Matlock (1999) and Gries (2002) argue that, although intuition may be poor as a methodology for investigating mental representations, linguists' intuitions are useful in the formulation of testable hypotheses about linguistic structure and behaviour.

Kepser and Reis (2005) point out that introspective and corpus data were the two main sources of data for theoretical linguistics until the mid-1990s. After that time other sources have been considered, such as experimentation (investigating offline and online processes), language acquisition, language pathologies, neurolinguistic, etc. They argue that linguistic evidence coming from different domains of data sheds more light on issues investigated than from a unique source. Multi-source evidence can either validate the theory or bring contradictory results, therefore opening new perspectives.

In what natural occurring data is concerned, Table 1 provides an overview of kinds of linguistic data (Gilquin and Gries 2009). They are presented in descending order of naturalness of production and collection (only corpora with written examples are produced for other aims than the specific purpose of linguistic research, and are thus the most natural kind).

In this chapter, we are interested in the first and the last type of data, namely corpora with written texts and data coming from experimentation where subjects are required to do something with language they do not usually do (using units they usually interact with involving typical linguistic output). We argue that both types

Table 1 Kinds of linguistic data (Sorted according to naturalness of production/collection) (Gilquin and Gries 2009: 5)

	Data source
1.	Corpora with written texts (e.g. newspapers, weblogs)
2.	Example collections
3.	Corpora of recorded spoken language in societies/communities where note-taking/recording is not particularly spectacular/invasive
4.	Corpora with recorded spoken language from fieldwork in societies/communities where note-taking/recording is spectacular/invasive
5.	Data from interviews (e.g. sociolinguistic interviews)
6.	Experimentation requiring subjects to do something with language they usually do anyway (e.g. sentence production as in answering questions in studies on priming or picture description in studies on information structure)
7.	Elicited data from fieldwork (e.g. response to "how do you say X in your language?")
8.	Experimentation requiring subjects to do something with language they usually do,
	*on units they usually interact with (e.g. sentence sorting, measurements of reaction times in lexical decision tasks, word associations)
9.	Experimentation requiring subjects to do something with language they usually do not do,
	*on units they usually interact with, involving typical linguistic output (e.g. measurements of event-related potentials evoked by viewing pictures, eye-movement during reading idioms, acceptability/grammaticality judgements
	*on units they usually do not interact with, involving production of linguistic output (e.g. phoneme monitoring, ultrasound tongue-position videos)

of data are complementary and necessary in pragmatic research, and may be used within various frameworks of linguistic description and analysis.

Before presenting the advantages and difficulties, as well as the complementarity of both empirical methods used in this study, we will define and describe briefly corpora and offline experiments.

2.1 Corpora

The well-known description of a corpus as being "a body of naturally occurring language" (McEnery et al. 2006: 4) is largely accepted in the corpus linguistics community, as well as other domain that work on corpora, such as empirical pragmatics or translation studies (Baker 1993, 1995). The same is true for corpora as having a *machine-readable* form, a feature that allows its compilation and analysis semi-automatically and automatically. As far as size is concerned, corpora become larger and larger and this is due to the possibility to be tagged, compiled and analysed automatically. The most important aspect to take into account when doing corpus work is to have an appropriate match of the research goal and the corpus type and size (Gries 2013).

Another feature of corpora is the number of languages and type of texts they contain, for example, monolingual or multilingual. Multilingual corpora can be of

two main types: (a) *parallel* (or *translation*) *corpora*, containing source texts and their translation in one or several target languages, which can be unidirectional (from language A to language B) or bi/multidirectional, and (b) *comparable corpora*, containing non-translated or translated texts of the same genre. Each type can be used for specific research goals.

A first advantage of working on corpora is that they represent an empirical basis for researchers' intuitions. Intuitions are the starting-point of any study but can be misleading and sometimes a few striking differences could lead to hazardous generalizations. Moreover, results of analyses of quantifiable data allow not only generalizations (through statistical significance tests) but also predictions through statistical analyses, such as correlations[3] or multiple regression models,[4] which are often used for investigating such a complex phenomenon as language.

Furthermore, multilingual corpora have quite naturally been used in contrastive studies. Contrastive Linguistics, also called Contrastive Analysis (CA), is "the systematic comparison of two or more languages, with the aim of describing their similarities and differences" (Johansson 2003: 31) and it is often done by focusing on one linguistic phenomenon. Mainly, the methodology used in a contrastive study consists of a first phase of monolingual description of the data (the phenomenon to be analysed), followed by the juxtaposition of two or more monolingual descriptions and the analysis of the elements according to a *tertium comparationis* (James 1980; Krzeszowski 1990). In our case study, we argue that the necessary *tertium comparationis* for verb tenses should be defined in terms of cross-linguistic valid features, such as *conceptual* and *procedural* information.

The practice of contrastive languages comparison based on corpora has itself numerous advantages, such as (a) new insights into the languages to be compared (which would have remained unnoticed in studies of monolingual corpora), (b) the highlighting of language-specific features and (c) the possibility of making semantic and pragmatic equivalences for the considered linguistic phenomenon between the SL and the TL. In some cases, corpus-based studies with a contrastive perspective have applicable purposes, such as our case study, which aims at modelling verb tenses for improving the quality of the texts translated by machine translation systems.

Another advantage is that data from corpora can be annotated (enriched) with semantic and pragmatic information, which allows more complex analyses. Annotation is the practice of adding interpretative linguistic information to a corpus, as underlined by Leech (2005). Annotation is thus an enrichment of the original raw

[3] Correlation is a monofactorial statistical method, which investigates the relation between one independent variable (the predictor) and one dependent variable (the phenomenon of interest). Correlation does not involve obligatorily causality between the two variables (they can be only associated) and can be used only when relationship is linear (cf. Gries 2009; Baayen 2008).

[4] Multiple regressions are multifactorial statistical methods, which investigate the relation between several independent variables (predictors) and one dependent variable, as well as their interactions. The relation between independent variables and the dependent variable can be linear or non-linear. (cf. Gries 2009; Baayen 2008).

corpus. From this perspective, adding annotations to a corpus is providing additional value and thus increasing their utility (McEnery and Wilson 2003; Leech 2004). Firstly, annotated corpora are useful both for the researcher(s) who made the annotation and for other researchers, who can use them for their own purposes, modify or enlarge them. Secondly, annotated corpora allow both manual and automatic analysis and processing of the corpus and by assuring its multifunctional utilisation, the annotations themselves often revealing a whole range of uses which would not have been practicable unless the corpus had been annotated. Thirdly, annotated corpora allow an objective record of analysis open to future analysis, decisions being more objective and reproducible. Due to automatic analysis of the corpus, annotated corpora are often used for training of NLP tools, such as classifiers (see Sect. 3.4).

Corpus work is thus interesting when the researcher is concerned with a descriptive approach of the linguistic phenomenon considered, as well as the study of language in use, given the fact that most of the time cotext and contextual information is also available in the corpus. Corpora permit monolingual and cross-linguistic investigations. Furthermore, corpus work allows the researcher to uncover on the one hand, what is probable and typical and, on the other hand, what is unusual about the phenomenon considered.

Corpus work has also some difficulties, such as the insufficiency of multilingual corpora for less widespread languages or the predilection for 'form-based research' where there is an interest in a specific grammatical form (Granger 2003). These difficulties constrain researchers to carry out their research manually, including building their corpus themselves and annotating it if they are interested in other phenomena than a specific grammatical form, such as semantic or syntactic categories. Another difficulty about corpus work is when the researcher is interested in infrequent phenomena[5] that will have insufficient occurrences in the corpus. Difficulties are also encountered when phenomena that are not lexically expressed such as world knowledge used in inferences as well as the cognitive basis of language are investigated.

This is one reason why corpus data are more and more combined with other types of evidence, such as experimentation. In what follows, we will briefly describe the use of experimentation in pragmatics and put forward the complementarity between corpus work and experimentation.

2.2 Experimentation

In pragmatics, experimentation is extremely useful for studying issues from the semantics/pragmatics interface and testing theories concerning the psychological

[5] For example, Grivaz (2012) who studied causality in certain pairs of verbs in a very large corpus and with human annotation experiments, found that less frequent pairs had a good causal correlation while very frequent pairs had a small causal correlation.

real competence native speakers have regarding semantics and pragmatics (Katsos and Breheny 2008).

One important distinction at the semantics/pragmatics interface was proposed by Grice (1975/1989) between what is 'said' vs. what is 'implicated' within the entire meaning of an utterance. The first experimental study of the identification and labelling by ordinary speakers of what is 'said' vs. what is 'implicated' was Gibbs and Moise (1997). In their chapter, Gibbs and Moise designed their experiments to determine whether people distinguished what speakers say from what they implicate and if they viewed what is 'said' as being enriched pragmatically. They used five categories of sentences[6] and participants had to choose between a minimal vs. enriched interpretation. Example (1) illustrates the *temporal relation* type of sentence as well as the two possible interpretations (minimal or literal meaning and the pragmatically enriched meaning):

(1) 'The old king died of heart attack and a republic was declared'.
(2) Minimal: order of events unspecified
(3) Enriched: the old kind died and then a republic was declared

The experiments were designed in order to manipulate the type of sentence, the instructions and the context of the targeted sentence. In the first experiment, the instructions consisted in explaining the two categories of interpretation of the sentence and no context was given. In the second experiment, the instructions were more detailed, including information about linguistic theories addressing the distinction between what is 'said' and what is 'implicated'. In the last two experiments, linguistic contexts were provided (a short story) in order to favour enriched interpretation (in the third experiment) as in example (4) and the minimal interpretation (in the fourth experiment) as in example (5), regarding *temporal relation* sentences.

(4) The professor was lecturing on the life of Jose Sebastian. He was a famous rebel in Spain who fought to overthrow the King. Many citizens wanted Sebastian to serve as their President. "Did Jose Sebastian ever became President?" one student asked. The professor replied, *The old king died of a heart attack before and a republic was declared.*
(5) Mike liked to take long bike rides each day. He also liked to sing as he rode because he has a terrific voice. Mike's roommate thought this was funny. He said to someone that *Mike likes to ride his bike and sing at the top of his lungs.*

Gibbs and Moise's four experiments showed that speakers assume that enriched pragmatics plays a significant role in what is said: the enriched interpretation was preferred in the first three experiments but not in the last one where the context biased strongly for the minimal interpretation. Manipulation of instructions and training did not have any effect on the participants' judgements.

[6] Cardinal (*Jane has three children)*, possession (*Robert broke a finger last night*), scalar (*Everyone went to Paris*), time-distance (*It will take us some time to get there*) and temporal relations.

We can make three observations concerning the temporal relation sentences: (a) temporal sequencing is an inference drawn contextually,[7] (b) it is independent of the specific instructions that speakers received and (c) it can be blocked in a context biasing for the minimal interpretation, that is the unspecified order. On the basis of their results, Gibbs and Moise argue that there might be two types of pragmatic processes, one that provides an interpretation for what speakers say and another one that provides an interpretation for what speakers implicate. They argue that this position can be explained by the principle of optimal relevance (Sperber and Wilson 1986/1995) and they acknowledge the difficulty of testing it experimentally. In our case study, we will consider temporal sequencing under the label [± narrativity] as being an inferential type of information that can function as a disambiguation criterion for usages of the English Simple Past (SP).

We now turn to experimentation as a type of methodology used in empirical and experimental pragmatics and we point out two advantages of adopting it: (a) it makes possible systematic control of confounding variables, and (b) depending on the nature of the experiment, it permits the study of online processes (Gilquin and Gries 2009: 9). One difficulty with experimentation is the artificial setting experiments require that can influence the behaviour of the participants in this unnatural setting. If experimental pragmatics completely adopted the psycholinguistics methodology as well as the study of online processes (through EEG and eye-tracking tools), empirical pragmatics focused mainly on offline experimentation, preserving the very essence of experimental studies: systematic manipulation of independent variables in order to determine their effect on dependent variables.

Concerning the complementarity of the two empirical sources of data, Gilquin and Gries argue that a corpus has a fourfold purpose in experimentation: (a) validator: the corpus serves as a validator of the experiment, (b) validatee: the corpus is validated by the experiment, (c) equal: corpus and experimental data are used on an equal footing and (d) stimulus composition: the corpus serves as a database for the items used in experiments. They also note that corpus work deals with a larger range of phenomena that can be investigated compared to experimentation. Experiments, however, allow the study of phenomena that are infrequent in corpora. Corpora and experiments have thus advantages and disadvantages that are complementary and thus linguists nowadays tend to use both of these empirical methods.

Finally, we would add that data from experiments represent human annotated data and can be used for NLP as training for automatic classifiers, thus proving the machines with different sorts of information (linguistic, contextual and world knowledge) that humans have and use in language interpretation process.

[7] In his Model of Directional Inferences (2000, 2002), Moeschler makes the same prediction about temporal relations between eventualities. They have an inferential nature and are drawn based on contextual assumptions. They can be blocked (minimal interpretation) under certain specific linguistic and contextual conditions.

In this research we consider data from experimentation (the 9th type of data in Gilquin and Gries'classification), focusing on linguistic judgments made by participants. Linguistic judgments were used mainly for acceptability and grammaticality tasks but nowadays they concern all types of linguistic information. By presenting our case study, we aim at pointing out the complementarity of corpus work and experimentation for testing theoretic hypothesis, build description models and apply them to NLP.

In what follows, we provide a case study presenting our investigation on verb tenses and show how the methodology presented above has been used, as well as how the results of our study support our thesis about the advantages of combining corpora work and experimentation when doing empirical pragmatics research.

3 Case Study

The case study presented in this article belongs to two research projects that aim[8] at improving the results of statistical machine translation (SMT) systems by modelling intersentential relations, such as those that depend on *verb tenses* and *connectives*. We investigate the 'meaning' of verb tenses, where the meaning is seen as consisting of both *what is said* and *what is implicated*. We deal thus with the semantics and pragmatics of verb tenses. Within the frame of empirical pragmatics, we study verb tenses within RT from a contrastive perspective based on parallel corpora and offline experimentation. Moreover, data from experimentation (human annotation) was used for automatic annotation and, furthermore, for training of a statistical machine translation (SMT) system.

As Aménos-Pons (2011) correctly underlines, any approach to tenses must deal with the fact that they present a certain stability of some basic features, combined with a high adaptability at discourse level that depends on contextual information (semantic and pragmatic) and world knowledge. A great challenge for linguists was, and remains, to know which of the features of verb tenses are stable and which are not.

Probably, one of the few generally accepted ideas about the meaning of verb tenses is the linguistic underdeterminacy thesis, as developed in RT and applied specifically to verb tenses by Neil Smith (1990). According to it, verb tenses are defined as a referential category: they can be characterized as locating temporal reference for eventualities with respect to three coordinates: speech moment S, event moment E and reference point R (Reichenbach 1947) through contextual enrichment following the expectation of optimal relevance (Wilson and Sperber 1998).

[8] The COMTIS Project (Improving the Coherence of Machine Translation Output by Modeling Intersentential Relations; project no. CRSI22_127510, March 2010-July 2013) and the MODERN Project (Modeling discourse entities and relations for coherent machine translation; project no. CRSII2_147653, August 2013–August 2016) belong to the Sinergia interdisciplinary program funded by the Swiss National Science Foundation.

The consequence of this theory is that verb tenses do not have several meanings but several usages corresponding to different contextual interpretations.

In the literature, two main trends are opposed regarding the nature of the encoded content verb tenses: on the one hand, verb tenses have only rigid procedural meanings that help the hearer reconstruct the intended representation of eventualities (Nicolle 1998; Aménos-Pons 2011; de Saussure 2003, 2011). de Saussure (2003) proposes algorithms to follow, consisting of the instructions encoded by verb tenses, in order to grasp the intended meaning of a verb tense at the discourse level.

On the other hand, verb tenses are seen as having both procedural and conceptual contents, as argued in Moeschler (2002) and Grisot et al. (2012). In Grisot et al. (2012) we argue that the conceptual content is given by a specific configuration of Reichenbachian coordinates event moment E, reference point R and speech moment S. The procedural content consists of instructions and constraints for contextual usages, namely [± narrative] and [± subjective]. Conceptual and procedural information represent bare-bone semantics that are contextually worked out through inferences (explicatures consisting of pragmatically determined aspects of what is said). The hearer has to ascertain the contextual value for both types of encoded information in order to access the right contextual hypotheses to get the intended cognitive effects.

Regards conceptual information, the assumption is that the specific configuration of the temporal coordinates S, R and E behaves like *pro-concepts* (Wilson 2011; Sperber and Wilson 1998: 15). *Pro-concepts* are semantically incomplete, they are conveyed in a given utterance and have to be contextually worked out. Once the enrichment process is completed the propositional form of the utterance is also available. This temporal information is not defeasible, i.e. it cannot be cancelled. The temporal coordinates S, R and E combine with the predicate's lexical aspect, in order to allow the calculation of the aspectual class (state, process, event). This conceptual information is the skeleton of the usage for each verb tense, which is enriched with contextual information and world knowledge in the inferential interpretation process.

Concerning the status of the temporal coordinates, de Saussure and Morency (2012) argue that tenses encode instructions on how the eventuality is to be represented by the hearer through the positions of temporal coordinates. They consider thus that temporal location with the help of S, R and E is of a procedural nature. We will show later on in this chapter that experimental studies revealed the contrary: the configuration of temporal coordinates is of a conceptual nature, specifically, they are variables that are saturated contextually.

The procedural content of verb tenses, on the other hand, consists of two types of instructions: (a) the [± narrative] instruction: to verify whether R is part of a series of points of reference available in the context and thus, eventualities are temporally sequenced, and (b) the [± subjective] instruction: to verify whether there is a perspective or a point of view on the eventuality presented. The experimental work that we conducted (see Sect. 3.3.3) showed that the [± narrative] feature includes temporal sequencing (inferential temporal relation as in Gibbs and

Moise's experiments described in Sect. 2.2) and causal relations holding between eventualities (cf. Moeschler 2003, 2011 for the relation between causality and temporal sequencing).

Another important point in the model described in Grisot et al. (2012) is that the specific combination of conceptual content and procedural content characterises contextual usages of verb tenses and not the meaning of a verb tense. For this point Grisot et al.'s analysis joins Aménos-Pons (2011) who assumes that tenses do not encode temporal relations. They are only the result of the tense meaning in specific environments.

In this chapter we adopt the view proposed by Grisot et al. (2012) and we bring new arguments, as well as evidence from experimental work, that support the procedural and conceptual nature of the information encoded by verb tenses expressing past time in French (FR) and English (EN).

3.1 Our Hypotheses

An investigation of parallel corpora consisting of several stylistic genres revealed the five most frequent translation divergences: (a) EN into French FR: the SP, the Simple Present and the Present Perfect (PresPerf), and (b) from FR into EN – the Passé Composé (PC) and Présent. In a first research phase, we chose to investigate the translation of the EN SP into FR, where its semantic and pragmatic domain is rendered through the Passé Simple (PS), the PC and the Imparfait (IMP). In order to grasp the meaning of the EN SP and its usages, we assume that the distinction between *conceptual* and *procedural* types of information is very important.

Our assumptions are: (1) a verb tense encodes conceptual and procedural information and (2) conceptual and procedural contents explain cross-linguistic variation. In what concerns the first hypothesis, we argue and bring evidence from offline experiments that procedural information encoded by the English SP is inaccessible to consciousness and hard to describe in conceptual terms, while conceptual information is accessible to conscious thinking and can be conceptualized. We also argue that the conceptual content of verb tenses (specifically, a specific configuration of temporal coordinates S, E and R) behaves like pro-concepts in that they are conveyed in a given utterance and have to be contextually worked out (explicature).

Concerning our second hypothesis, we assume that conceptual and procedural contents of verb tenses explain their cross-linguistic variation revealed by an investigation of our parallel corpora. A verb tense can have several usages, where each usage is triggered by a language-specific combination of conceptual and procedural contents. Parallel corpus analysis reveals that each usage of a verb tense in a SL is rendered by a different verb tense in a TL. Specifically, the translation divergence of the English SP into FR can be resolved if contextual usages of the SP are considered.

In the following sections, we bring evidence for our model for the semantics and pragmatics of the English SP from parallel corpus (Sect. 3.2) and offline experiments (Sect. 3.3). Section 3.4 is dedicated to the NLP application of the model defended in this case study.

3.2 Data from Parallel Corpora with a Contrastive Perspective

In Grisot and Cartoni (2012) we studied the discrepancies between theoretical descriptions of verb tenses and their use in parallel corpora. We investigated corpora consisting of texts in EN and their translations into FR that belong to four different genres (literature 18 %, journalistic 18 %, legislation 33 % and EuroParl 31 %). A total of 1275 predicative verb tenses have been considered, which represents 77 % of the verb tenses occurring in the corpus. The qualitative and quantitative analysis of the corpus was done in two steps. In the first monolingual step, we identified tenses that occur in the corpus and calculated their frequency in the SL. In the second bilingual step, we identified the tenses used as translation possibilities in the TL of a certain tense from SL and calculated their frequency. Analysis of frequency of tenses in SL provided information about tenses that are possible candidates for being problematic for machine translation systems. The asumption is that frequent tenses, if wrongly translated, decrease the quality of the translated text. Bilingual analysis with focus on identifying verb tenses used as translation possibilities in TL for ambiguous tenses in SL revealed that the SP is translated into FR using mainly three tenses (PS, PC and IMP representing 80 % of translation possibilities) as in examples (6), (7) and (8) and that the PresPerf is translated using two tenses (PC and Présent, 100 % of translation possibilities) as in examples (9) and (10). These are two of the translation divergences shown by analysis of parallel corpora.

(6) EN/SP: General Musharraf **appeared** on the national scene on October 12, 1999, when he ousted an elected government and announced an ambitious "nation-building" project. (Journalistic Corpus: "News Commentaries")

FR/PC: Le Général Moucharraf **est apparu** sur la scène nationale le 12 octobre 1999, lorsqu'il a forcé le gouvernement élu à démissionner et annoncé son projet ambitieux de "construction d'une nation".

(7) EN/SP: With significant assistance from the United States—warmly accepted by both countries—disarmament was orderly, open and fast. Nuclear warheads **were returned** to Russia. (Journalistic Corpus: "The New York Times")

FR/PS: Avec l'assistance non négligeable des Etats-Unis – chaleureusement acceptée par les deux pays: le désarmement a été méthodique, ouvert et rapide. Les ogives **nucléaires furent renvoyées** en Russie.

(8) EN/SP: He **seemed** about seventeen years of age, and was of quite extraordinary personal beauty, though somewhat effeminate. (Literature Corpus: O. Wilde, "The picture of Mr. W.H")

(continued)

FR/IMP: Il **paraissait** avoir seize ans, et il était d'une beauté absolument extraordinaire, quoique manifestement un peu efféminée.

(9) EN/PresPerf: I would like to fully support Mrs Roth-Behrendt's proposals, but we have spent over 20 years talking about people's willingness to spend more money on food; it is just that the distribution process **has** totally **changed**. ("EuroParl" Corpus)

FR/Présent: Je soutiendrais vraiment de tout coeur les propositions de Mme Roth-Behrendt; cela fait vingt ans que nous parlons de la possibilité de consacrer plus d' argent à l' alimentation mais, quand il s' agit du processus de distribution, c'**est** tout autre chose.

(10) EN/PresPerf: Whether or not the government was involved, the fact remains that Pakistan **has lost** a desperately needed leader. (Journalistic Corpus: "News Commentaries")

FR/PC: Que le gouvernement soit ou non impliqué, le fait est que le Pakistan **a perdu** un leader dont il a cruellement besoin.

The ambiguity of the EN SP, as well as the PresPerf, is illustrated by their translation into FR. In order to improve their translation by SMT systems, these tenses must be disambiguated. Following the CA's methodology, the SP and the PresPerf, as well as the FR tenses used for their translation, must be compared in three steps. The first step consists of the monolingual description, followed by bilingual juxtaposition of the two monolingual descriptions and finally, their analysis according to the *tertium comparationis* defined in terms of conceptual and procedural contents.

Now in what concerns the SP, known as *preterit*, it describes an action or state as having occurred or having existed at a past moment or during a past period of time that is definitely separated from the actual present moment of speaking or writing. Comrie (1985: 41) emphasized that the SP "only locates the event in the past, without saying anything about whether the situation continues up to the present or into the future". Radden and Dirven (2007: 219) argue that the use of the SP to express bounded past situations, presented as a series of events, typically in narratives, as in (11). The individual events from example (11) are temporally ordered (signalled by the coordination and the conjunction *and*) and are thus interpreted as being successive.

(11) I *grabbed* his arm and I *twisted* it up behind his back and when I *let* go his arm there *was* a knife on the table and he just *picked* it up and *let* me have it and I *started* bleeding like a pig. (Labov and Waletzky 1967, quoted by Radden and Dirven 2007: 219)

The most frequent verb tenses used in FR for translating the SP are, as we have already noted, the PC, PS and IMP. The PC is classically described from a monolingual point of view as a "tense with two faces" (Martin 1971) because it can express both past and present time. The PS is described as a tense that expresses a past event completely accomplished in the past with no connection to present time (Grevisse 1980, Wagner and Pinchon 1962) and used in contexts where events are temporally ordered (Kamp and Rohrer 1983). Finally, the IMP is a tense that expresses background information (Weinrich 1973). The focus on the accomplishment of the event in the past is the feature that distinguishes the PS from the PC, the second one expressing a link to present time, while perfectivity is a feature that distinguishes the PS from the IMP, the former being perfective and the latter imperfective.

Given these monolingual descriptions, when juxtaposed, we can observe the multitude of facets for describing these four tenses: in terms of temporal location (time preceding, simultaneous or even following speech moment), grammatical aspect (perfective or imperfective), discursive grounding (foreground or background information) and relation to other eventualities (temporally ordered or not). Another point that can be observed is the lack of a one-to-one correspondence between the several meanings of the SP and the three FR tenses used for its translation. In Grisot et al. (2012), we argue that the meaning of these verb tenses should be investigated cross-linguistically in terms of their conceptual and procedural information, and more specifically that the procedural information [$\pm$ narrativity] is a disambiguation criterion for the usages of the SP. In this study we bring evidence for our claim that the [$\pm$ narrativity] feature is procedural (through experimental work presented in Sect. 3.3.3). We show that occurrences of SP annotated by two human annotators as having a narrative usage correspond in the parallel corpora investigated to translation through either PS or PC and occurrences annotated as having a non-narrative usage correspond to translation through an IMP (detailed results provided in Sect. 3.3.3).

The EN PresPerf is characterized by a grammatical combination of present tense and perfect aspect and it is used to express a past eventuality that has present relevance. The same grammatical combination exists in other languages such as the FR PC, with the specificity that the PC can also express eventualities accomplished in the past. In EN, there is a competition between the SP and the PresPerf for referring to past time eventualities, with the particularity that PresPerf is incompatible with adverbials expressing define past time. The first annotation experiment considered the competition between SP and PresPerf forms for expressing past time eventualities, showing that each verb tense encodes conceptual information and it can easily be dealt with by human annotators (Sect. 3.3.2).

A benefit of parallel corpora is the availability of context and cotext, information that facilitates establishing semantic and pragmatic equivalence for each verb tense. This information is crucial as regards usages of verb tenses.

From the corpus described above, we used a subset of 30 excerpts randomly selected (that we call *items* and all contain occurrences of the SP or PresPerf) for the first experiment and 458 items (containing occurrences of the SP) for the second experiment. In what follows, we describe and provide the results of annotation experiments.

3.3 Data from Offline Experiments

Experimental work we have conducted brought evidence for the hypothesis that verb tenses encode both conceptual and procedural information. Conceptual information concerns different combinations of Reichenbachian temporal coordinates, which are contextually saturated variables. Procedural information concerns instructions relating temporal and causal relations holding between the eventualities expressed

in the sentence. In this section, we will provide the general design of our experiments (participants, procedure and evaluation), followed by the presentation of the two experiments and their results.

3.3.1 Design of Experiments and Participants

The two annotators were native speakers of EN with basic knowledge of FR. They were asked to follow the instructions (given below for each type of information annotated) and went through a training phase in order to check whether the instructions given were clear and correctly understood. For the effective annotation task, annotators received a file with the total number of excerpts that were taken from the EN part of the parallel corpora. For each item, sentences including the verb tense considered, as well as one sentence before or after, were provided in order to have sufficient context for pragmatic judgement.

One way of evaluating human annotation is to calculate the inter-annotator agreement with the help of the *kappa* coefficient (Carletta 1996). One issue that influences corpus annotation by raters is the subjectivity of the judgements, which can be quite substantial for semantic and pragmatic annotations (Artstein and Poesio 2008). It can be tested whether different raters produced consistently similar results, so that one can infer that the annotators have understood the guidelines and that there was no agreement just by chance. The kappa statistic factors out agreement by chance and measures the effective agreement by two or more raters. The kappa coefficient has values between 0 to 1, going from no agreement other than that expected to occur by chance to total agreement among raters. We used this measure for quantifying the inter-annotator agreement in our experiments.

3.3.2 Annotation of Conceptual Information

Through this annotation experiment, we wanted to determine the conceptual meaning of two verb tenses in EN, SP and PresPerf. Our expectation was that human annotators should be able to think of the meaning of SP and PresPerf consciously, conceptualize it and make specific decisions in each context with easiness. Annotators received annotation guidelines (presented below) and went through a training phase before the actual annotation phase.

As there are no quantitative measures[9] proposed in the literature to evaluate the conceptual and procedural type of information encoded by linguistic expressions, at least none that we are aware of, we propose to use the kappa coefficient to quantify

[9] de Saussure (2011) proposes a qualitative criterion to evaluate procedural expressions: an expression is procedural if it triggers inferences that cannot be predicted on the basis of an identifiable conceptual core to which general pragmatic inferential principles are identified.

conceptual and procedural information. Wilson and Sperber (1993) and Wilson (2011: 11) descibe conceptual information as accessible to consciousness, capable of being reflected on, evaluated and used in general inference, and procedures as "relatively inaccessible to consciousness, resistant to conceptualisation, thus we can not discover through introspection the rules of our language, the principles governing inferential comprehension, or the processes involved in mental-state attribution". We assumed thus that manipulating conceptual information described as easily graspable concepts is related to the notions of sensitivity and accessibility to consciousness, specifically native speakers' sensitivity is a cue to direct access to the encoded conceptual content. We expected thus high values of the inter-annotator agreement coefficient based on the relative facilty of the task, namely to identify striking information.

As far as procedural information is concerned, we expected low agreement, related to a more difficult task: procedural information is notoriously hard to pin down in conceptual terms (Wilson and Sperber 1993:16) and not accessible to consciousness. The processing of the narrative feature is predicted to be less accessible because it is the result of a non-guaranteed pragmatic inference (*non-demonstrative inference*[10] for Sperber and Wilson 1986/1995: 65) based on conceptual information, cotextual information and contextual hypotheses. As inferential processes are costly and depend on several factors, they are predicted to produce lower values of the inter-annotators agreement coefficient.

Based on our claim (Grisot et al. 2012) that the configuration of Reichenbachian coordinates should be split into three pairs of two coordinates (E/R, R/S and the inferred E/S) instead of the classical view of three coordinates as Reichenbach proposed. We defined the conceptual content of the Simple Past, as in example (12) to be the pair $E < S$ which bears the focus (from the line $E = R$, $R < S$ and $E < S$), in other words '*situation that happened in the past*' and the conceptual meaning of Present Perfect, as en example (13) to be the pair $R = S$ (from the line $E < R$, $R = S$, $E < S$),[11] in other words the "*current resulting state of a past situation*".

(12) EN/SP: After almost a decade in self-imposed exile, Bhuto's return to Pakistan in October **gave** her a fresh political start. Pakistan had changed, as military dictatorship and religious extremism in the north played havoc with the fabric of society. (Journalistic Corpus: "NewsCommentaries")

(13) EN/PresPerf: Some of the proposals concerning greater focus on equality **have** also **been accepted**, but the Council did not want to accept some very central proposals from Parliament. ("EuroParl" Corpus)

[10] Sperber and Wilson (1986/1995: 65) argue that the process of inferential communication is non-demonstrative: even under the best circumstances, it might fail (the addressee can not deduce the communicator's communicative intention).

[11] In the parallel corpus both the SP and the PresPerf from these two examples are translated by a PC in French, highlighting thus another translation divergence: the French PC into EN. A hint of the disambiguation criterion is a focus either in the $E < S$ relation for the SP or on the $R = S$ relation for the PresPerf (as we argued in Grisot et al. 2012).

The annotation guidelines included: (a) a description of the two types of meaning (b) one example for each usage, as given in the examples below and (c) the instruction to read each excerpt, identify the meaning of the verb highlighted and decide on the type of usage. In the first example, the most salient information is the result state in the present: the fact that the false declaration is now filled. In the second example, the most salient information is the situation that happened in the past: the lack of choice of Musharraf.

(14) And instead of full cooperation and transparency, Iraq **has filed** a false declaration to the United Nations that amounts to a 12,200-page lie. (Journalistic Corpus: "NewsCommentaries")

(15) In a historic ruling that Musharraf **had** little choice but to accept, the Supreme Court itself reinstated the Chief Justice in July. Subsequently, the energized judiciary continued ruling against government decisions, embarrassing the government – especially its intelligence agencies. (Journalistic Corpus: "NewsCommentaries")

In what concerns the annotation guidelines, three aspects should be mentioned: (a) the 'meaning' of the SP and PresPerf, respectively, was easily identified and conceptualized in order to explain the task to annotators, (b) they were asked to identify 'the most salient information' in order to identify the focus and (c) annotators understood the annotation task easily, as well as the examples used for training.

In this experiment, annotators made decisions on 30 excerpts from the corpus following the annotation instructions. They agreed on all the items annotated (kappa = 1) and pointed out the easiness of the task. This result can be interpreted as evidence for the conceptual nature of the information considered in this experiment. We assume that the total agreement is due to the highly accessible conceptual information, that is, the ability for the raters to consciously represent the temporal coordinates as part of the conceptual meaning of tenses.

3.3.3 Annotation of Procedural Information

One of the features tested with the help of the annotation experiment is [± narrativity]. As mentioned, this feature is a procedural information encoded by tenses that instructs the hearer/reader to verify whether the reference point is part of a series of R that increases incrementally, in other words if the eventualities presented are temporally ordered. Wilson (2011) emphasized that procedures are not part of the meaning of a linguistic expression but are merely activated or triggered by the occurrence of that expression in an utterance. If the feature is activated ([+ narrative]), then we can talk about a *narrative* usage of the verb tense considered. And respectively, if the feature is not activated [non-narrative], then the verb tense considered has a non-narrative usage.

Numerous studies have already addressed narrativity either in the traditional rhetoric (since the nineteenth century, such as Alexander Bain 1866 and John Genung 1900), in DRT (Kamp and Reyle 1993) and SDRT (Lascarides and Asher 1993) or within a semantics and pragmatics perspective (Hinrichs 1986; Partee 1984; Reboul and Moeschler 1998; Smith 2001, 2003, 2010). Mainly, in these studies, narrativity is a discourse relation or a discourse mode associated with temporal sequencing of eventualities. In this chapter, we adopt this view of narrativity and postulate that it is a binary variable ([± narrativity]) that represents procedural information conveyed by verb tenses and which can be used as a disambiguation criterion for various usages of tenses expressing past time in EN and FR.

The verb tense considered in this annotation experiment is the EN SP. As in the first experiment, annotators received annotation guidelines (presented below) and went through a training phase. Narrativity was defined and explained to annotators as it follows:

(16) In *narrative* contexts a story that is being told (you might not have the whole story available in the sentence) and eventualities are temporally ordered, while *non-narrative* contexts are associated with descriptive passages, where no story is being told.

Annotation guidelines included: (a) a definition of narrativity (b) the explanation of each usage (narrative and non-narrative) with two examples for each usage, as given in the examples below, (c) the instruction to read each excerpt, identify the verb highlighted and decide if in context, the highlighted verb is part of the underlying theme (the verb tense would have a narrative usage) or not (the verb tense would have a non-narrative usage).

In the first example below, there are two events, i.e. 'the marriage that happened' and 'the wealth which was added'. The second event is presented in relation to the first (first he got married and then he added to his wealth), which is why the SP verbs happened and added are in narrative usage. In the second example, there are three states (was a single man, lived and had a companion) that describe the owner of the estate. States are not temporally ordered, which is why this example illustrates the non-narrative usage of the SP.

(17) By his own marriage, likewise, which happened soon afterwards, he **added** to his wealth. (Literature Corpus: J. Austen, "Sense and Sensibility")

(18) The late owner of this estate was a single man, who **lived** to a very advanced age, and who for many years of his life, had a constant companion and housekeeper in his sister. (Literature Corpus: J. Austen, "Sense and Sensibility")

The value of kappa coefficient for this annotation experiment was 0.42. This value is above chance, but not high enough to point to entirely reliable linguistic decisions (values generally accepted around 0.6–0.7). What this first result shows about the procedural feature [± narrativity] encoded by the EN SP is the difficulty hearers/readers have in the interpretation process to conceptualize the language rules they have and make decisions about their functioning.

The two annotators agreed on 325 items (71 %) and disagreed on 133 items (29 %). Error analysis showed that the main source of errors was the length of the temporal interval between two eventualities, which was perceived differently by the two annotators. This lead to ambiguity between temporal sequence or simultaneity, each of them corresponding to narrative, respectively, non-narrative usage, as in example (19) where the eventualities "qualify" and "enable" were perceived as being simultaneous by one annotator and successive by the other.

(19) Elinor, this eldest daughter, whose advice was so effectual, possessed a strength of understanding, and coolness of judgment, which qualified her, though only nineteen, to be the counsellor of her mother, and **enabled** her frequently to counteract, to the advantage of them all, that eagerness of mind in Mrs. Dashwood which must generally have led to imprudence. (Literature Corpus: J. Austen, "Sense and Sensibility")

A possible explanation is the fact that personal world knowledge is used to infer temporal information, such as the length of the temporal interval between two eventualities, i.e. information that allows the annotator to decide whether the eventualities are temporally ordered or not. Cases where the length of the temporal interval between two eventualities was very reduced were ambiguous for the annotators, so each of them decided differently whether it was long enough for temporal sequencing or too short, so that the simultaneity meaning was preferred.

Disagreements were resolved in a second round of the annotation experiment, where the narrativity feature was identified with a new linguistic test that was explained to two new annotators.[12] Annotators were asked to insert a connective such as *and* and *and then* when possible, in order to make explicit the 'meaning' of the excerpt, namely the temporal relation existent between the two eventualities considered. The connective *because* (for a causal relation) has also been proposed by annotators under the [+ narrative] label showing that causal relations should also be considered. We thus considered causal relations under the [+ narrative] tag but we will not look more into causality in this chapter. The inter-annotator agreement in this second experiment was *kappa* = 0.91, signalling very strong and reliable agreement. This result emphasizes the procedural nature of the feature taking into account that one of the characteristics is the possibility to render explicit the instructions encoded with the help of discourse markers.

The cross-linguistic application of these findings consists of the observation of a pattern in the parallel corpus. We investigated the data containing agreements from both annotation rounds (435 items) and analyzed them in the parallel corpus. We observed that the narrative usages of the SP identified by annotators correspond to narrative usages[13] in the FR part of the corpus (translation by a PC or PS) and the

[12] The new annotators were one of the authors and a research peer, who was not aware of the purpose of the research.

[13] In Grisot et al. (2012), we describe a similar annotation experiment made on the French tenses used for translating the EN SP, namely PC, PS and IMP. In this experiment, the PC and PS have been identified as being narrative and the IMP as being non-narrative with a kappa value of 0.63 (reliable agreement).

non-narrative usages of the SP correspond to the non-narrative usages in the FR text (translation with an IMP) in 338 items (78 %). This leaves 22 % where annotators agreed on the narrativity label but where it is not consistent with the tense used in FR. Future work will focus on investigating the other factors that explain the 22 % of the variation in the translation of the SP in French.

3.4 Natural Language Processing Application

Nowadays, linguistic research tends more and more to integrate language automatic processing techniques. Human annotation and classification of texts is often used in NLP and Machine Translation (MT). Most of the current MT systems incorporate a language model that analyses texts at the sentence level. But there are linguistic phenomena whose interpretation is done using information that goes beyond sentence boundaries, such as verb tenses. The theoretical model of the pragmatics and semantics of the EN SP described in this chapter has been validated empirically also through an NLP technique called *automatic annotation* or *classification*. Human-annotated data provides to the machine translation system pragmatic information that humans make use of in the interpretation process, such as the reference point R, the relative sequence of eventualities, the length of the interval and any causal relation existent between eventualities.

Human-annotated texts described in this chapter served as training data for machine-learning tools,[14] specifically a maximum entropy classifier (Manning and Klein 2003). A classifier is a machine-learning tool that will take data items and place them into one of the available classes (in the present case, narrative and non-narrative) according to a statistical algorithm. The underlying principle of maximum entropy is that, when assigning a class, it should be done uniformly (uniform distributions) unless there is some external knowledge that would instruct the system to do it differently. Annotated data used for training these classifiers provide external knowledge and thus inform the automatic labelling technique where to be minimally non-uniform. Iterative runs of the classifier results in automatically labelled or annotated texts with the considered features.

The feature tested in our case study was [± narrativity] and the human-annotated data was used for training the classifier (see Grisot and Meyer 2014). The results of automatic annotation are similar to human annotation; the classifier correctly annotated 76 % of the items. The purpose of using automatic annotation is the possibility to do it on large amounts of data. Human annotation has the disadvantages of being tedious and costly, and it is often done on a reduced amount of data.

The final purpose was to improve the results in what concerns verb tenses of a statistical machine translation system. Current machine translation systems have

[14] The NLP work was done by our colleagues Thomas Meyer and Andrei-Popescu Belis from the Idiap Research Institute (Martigny, Switzerland) to whom we address our gratitude.

difficulties in choosing the correct verb tense translations, in some language pairs, because these depend on a larger context than systems consider. A machine translation system generally misses information from previously translated sentences, which is detrimental to lexical cohesion and coherence of the translated text.

A first run of an SMT system, which uses the classifier trained on the annotated data with the [± narrativity] feature, had slightly better results than without this pragmatic feature. When trained and tested on automatically annotated data, the [± narrativity] feature improves translation by about 0.2 BLEU points.[15] More importantly, manual evaluation shows that verb tense translation and verb choice are improved by respectively 9.7 % and 3.4 % (absolute), leading to an overall improvement of verb translation of 17 % (relative) (for more detailed results see Meyer et al. 2013).

4 Conclusion

This chapter has given an account of the place of empirical pragmatics among theoretical pragmatics and experimental pragmatics, for the study of language in use. We have argued for the need to have robust data for pragmatic research, data provided by both corpus work and experimentation.

We have shown that corpus work can be fruitfully done with a contrastive perspective, following the specific three-steps methodology of CA. As far as experimentation is concerned, we have looked into offline experiments consisting of linguistic judgement task that resulted in human annotated data. We have discussed the example of the first experiment for the pragmatic distinction between what is 'said' and what is 'implicated' designed by Gibbs and Moise (1997). Another important topic of this chapter was the discussion about the advantages and difficulties of each of the two methods considered (corpus work and experimentation), as well as their complementarity.

In our case study, we investigated the nature of the information encoded by verb tenses. We assumed and validated empirically through annotation experiments that verb tenses encode both procedural and conceptual information. We defined conceptual information as being involved in the language of thought in a Fodorian framework (Fodor 1975, 1998) having the characteristic of being accessible to consciousness and capable of being reflected on, evaluated and used in general inference. We proposed thus, based on these two features, that verb tenses encode conceptual information consisting of a certain configuration of temporal coordinates. The basic meaning of a tense is to locate an eventuality related to the speech moment, passing through a reference point. A verb tense encodes instructions to verify the

[15] BLEU (Bilingual Evaluation Understudy) is an evaluation measure for machine-translated texts. It calculates the degree of resemblance to a human-translated text and it is a number between 0 and 1, where values closer to 1 represent more similar texts.

contextual value of several features that are important and relevant for utterance comprehension. In this chapter, we investigate one feature: [± narrativity].

As far as procedural information is concerned, we followed Wilson and Sperber's idea (1993) that procedures are not part of language of thought and thus are not accessible to consciousness and easily conceptualized, as representations are. The results of the annotation experiment showed that verb tenses encode procedural information that instruct the reader/hearer to look for other eventualities that are related to the eventuality considered, namely the [± narrativity] procedural feature.

Taken together, the empirical findings of this research provide an example of the relation between theoretical framework(s) and empirical methodologies. Theoretical hypotheses have an impact on the choice of empirical methodologies. For example, a cross-linguistic perspective requires work on parallel corpora in order to have access to both source and target texts. The disambiguation of the usages of the targeted verb tense requires the formulation of possible disambiguation criteria that need to be validated through experimentation involving linguistic judgement tasks. Genuine data dealt with empirical methods can challenge theoretical positions. For verb tenses, for example, the results of our experiments challenged the theoretical assumption that verb tenses do not encode conceptual information, but only procedural information. Next to existent qualitative measures for conceptual and procedural information, we proposed a quantitative measure: the kappa coefficient for inter-annotator agreement. This measure makes use of the knowledge that native speakers have about their language.

Finally, our work has illustrated how empirical pragmatics can work together with the NLP domain. The pragmatic feature identified as procedural information and validated through human annotation experiments has been used as a label for discourse tagging with an automatic classifier. Moreover, a SMT system trained on the annotated corpus had better results for translating verb tenses than if it hadn't made use of the [± narrativity] pragmatic feature.

An issue that was not addressed in this study was the cross-linguistic application of the model to more than one pair of languages. This issue will be addressed in further studies and it targets the translation of the English SP into Italian and Romanian. The application of the conceptual/procedural distinction for verb tenses could also be done using online experimental methodology. This would probably reduce any remaining doubts about the existence of a conceptual content of verb tenses.

References

Amenós-Pons, J. (2011). Cross-linguistic variation in procedural expressions: Semantics and pragmatics. In V. Escandell-Vidal, M. Leonetti, & A. Ahern (Eds.), *Procedural meaning: Problems and perspectives* (pp. 235–266). Bingley: Emerald Group Publishing Limited.

Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics, 34*, 555–596.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
Bain, A. (1866). *An English grammar*. London: Longman, Green, Longman, Roberts, and Green.
Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 233–250). Amsterdam: John Benjamins.
Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target, 7*, 223–243.
Barlow, M., & Kemmer, S. (2000). Introduction: A usage-based conception of language. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 1–63). Stanford: Center for the Study of Language and Information Publications.
Blakemore, D. (1987). *Semantic constraints on relevance*. Oxford: Blackwell.
Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press.
Boas, H. C. (2003). *A constructional approach to resultatives*. Stanford: Center for the Study of Language and Information Publications.
Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics, 22*, 249–254.
Carston, R. (2002). Linguistic meaning, communicated meaning and cognitive pragmatics. *Mind & Language, 17*, 127–148.
Comrie, B. (1985). *Tense*. Cambridge: Cambridge University Press.
de Saussure, L. (2003). *Temps et pertinence: éléments de pragmatique cognitive du temps*. Bruxelles: De Boeck.
de Saussure, L. (2011). On some methodological issues in the conceptual/procedural distinction. In V. Escandell-Vidal, M. Leonetti, & A. Ahern (Eds.), *Procedural meaning: Problems and perspectives* (pp. 55–79). Bingley: Emerald Group Publishing Limited.
de Saussure, L., & Morency, P. (2012). A cognitive-pragmatic view of the French epistemic future. *Journal of French Language Studies, 22*, 207–223.
Escandell-Vidal, V., Leonetti, M., & Ahern, A. (Eds.). (2011). *Procedural meaning: Problems and perspectives*. Bingley: Emerald Group Publishing.
Fodor, J. A. (1975). *The language and thought*. Oxford: Harvard University Press.
Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Cambridge, MA: Oxford University Press.
Gazdar, G. (1979). *Pragmatics: Implicature, presupposition, and logical form*. New York: Academic.
Genung, J. F. (1900). *The working principles of rhetoric: Examined in their literary relations and illustrated with examples*. Boston: Ginn & Company.
Gibbs, R. W., & Matlock, T. (1999). Psycholinguistics and mental representations. *Cognitive Linguistics, 10*, 263–270.
Gibbs, R. W., & Moise, J. F. (1997). Pragmatics in understanding what is said. *Cognition, 62*, 51–74.
Gilquin, G., & Gries, S. T. (2009). Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory, 5*, 1–26.
Granger, S. (2003). The corpus approach: A common way forward for contrastive linguistics and translation studies. In S. Granger, J. Lerot, & S. Petch-Tyson (Eds.), *Corpus-based approaches to contrastive linguistics and translation studies* (pp. 17–29). Amsterdam: Rodopi.
Grevisse, M. (1980). *Le bon usage: grammaire française avec des remarques sur la langue française d'aujourd'hui* (11th ed.). Paris/Gembloux: Duculot.
Grice, H. P. (1975/1989). *Logic and conversation*. Harvard: Harvard University Press.
Gries, S. T. (2002). Evidence in linguistics: Three approaches to genitives in English. *The LACUS Forum, 28*, 17–31.
Gries, S. T. (2009). *Quantitative corpus linguistics with R: A practical introduction*. New York: Routledge.
Gries, S. T. (2013). *Statistics for linguistics using R* (2nd ed.). Berlin: De Gruyter Mouton.

Grisot, C., & Cartoni, B. (2012). Une description bilingue des temps verbaux: étude contrastive en corpus. *Nouveaux Cahiers de linguistique française, 30*, 101–117.

Grisot, C., & Meyer, T. (2014). *Cross-linguistic manual and automatic annotation for a pragmatic feature of verb tense*. Proceedings of LREC 2014, Reykjavik.

Grisot, C., Cartoni, B., & Moeschler, J. (2012). *Conceptual and procedural information for verb tense disambiguation: The English simple past*. Oral presentation at *EPICSV*, Seville, Spain.

Grivaz, C. (2012). *Automatic extraction of causal knowledge from natural language texts*. PhD thesis. Université de Genève, Geneva.

Hinrichs, E. (1986). Temporal anaphora in discourses of English. *Linguistics and Philosophy, 9*, 63–82.

Horn, L. R. (1973). Greek Grice: A brief survey of proto-conversational rules in the history of logic. *CLS, 9*, 205–214.

Horn, L. R. (1984). A new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form and use in context* (pp. 11–42). Washington, DC: Georgetown University Press.

Horn, L. R. (1989). *A natural history of negation*. Chicago: University of Chicago Press.

Horn, L. R. (1992). The said and the unsaid. *Ohio State University Working Papers in Linguistics, 40*, 163–192.

Horn, L. (2004). Implicature. In L. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 3–28). Oxford: Blackwell.

Horn, L. R. (2007). Toward a Fregean pragmatics: Voraussetzung, Nebengedanke, Andeutung. In I. Kecskes & L. R. Horn (Eds.), *Explorations in pragmatics. Linguistic, cognitive and intercultural aspects* (pp. 39–69). Berlin: Mouton de Gruyter.

James, C. (1980). *Contrastive analysis*. Burnt Mill: Longman.

Johansson, S. (2003). Contrastive linguistics and corpora. In S. Granger, J. Lerot, & S. Petch-Tyson (Eds.), *Corpus-based approaches to contrastive linguistics and translation studies* (pp. 31–45). Amsterdam: Rodopi.

Kamp, H., & Rohrer, C. (1983). Tense in texts. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, use and interpretation of language* (pp. 250–269). Berlin: Walter de Gruyter.

Kamp, H., & Reyle, U. (1993). *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation*. Dordrecht: Kluwer Academic Publishers.

Katsos, N., & Breheny, R. (2008). Two experiments and some suggestions on the meaning of scalars and numerals. In N. T. Enikó & K. Bibok (Eds.), *The role of data at the semantics-pragmatics interface* (pp. 125–160). Berlin: Mouton de Gruyter.

Katsos, N., & Cummins, C. (2010). Pragmatics: From theory to experiment and back again. *Language and Linguistics Compass, 4*, 282–295.

Kepser, S., & Reis, M. (2005). *Linguistic evidence: Empirical, theoretical and computational perspectives*. Berlin: Mouton de Gruyter.

Krzeszowski, T. P. (1990). *Contrasting languages: The scope of contrastive linguistics*. Berlin: Mouton de Gruyter.

Labov, W., & Waletzky, J. (1967). Narrative analysis: Oral versions of personal experience. In J. Helm (Ed.), *Essays on the verbal and visual arts: Proceedings of the 1966 annual spring meeting of the American Ethnological Society* (pp. 12–44). Seattle: University of Washington Press.

Lascarides, A., & Asher, N. (1993). Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and Philosophy, 16*, 437–493.

Leech, G. N. (2004). *Meaning and the English verb*. Edinburgh: Pearson Education.

Leech, G. N. (2005). Adding linguistic annotation. In W. Martin (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 17–29). Oxford: Oxbow Books.

Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Manning, C., & Klein, D. (2003). Optimization, MaxEnt models, and conditional estimation without magic. In *Tutorial at HLT-NAACL and 41st ACL conferences*, Edmonton, Canada and Sapporo, Japan.

Martin, R. (1971). *Temps et aspect: essai sur l'emploi des temps narratifs en moyen francais*. Paris: Klincksieck.
McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh: Edinburgh University Press.
McEnery, T., & Wilson, A. (2003). *Corpus linguistics*. Edinburgh: Edinburgh University Press.
McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies*. London/New York: Routledge.
Meibauer, J., & Steinbach, M. (2011). *Experimental pragmatics/semantics*. Amsterdam: John Benjamins.
Meyer, T., Grisot, C.,& Popescu-Belis, A. (2013). *Detecting narrativity to improve English to French translation of simple past verbs*. http://www.idiap.ch/~tmeyer/res/Meyer-DISCOMT-3-2013.pdf. Accessed 11 Sept 2013
Moeschler, J. (1989). Pragmatic connectives, argumentative coherence, and relevance. *Argumentation, 3*(3), 321–339.
Moeschler, J. (2000). Le modèle des inférences directionnelles. *Cahiers de linguistique française, 22*, 57–100.
Moeschler, J. (2002). Connecteurs, encodage conceptuel et encodage procédural. *Cahiers de linguistique française, 24*, 265–292.
Moeschler, J. (2003). Causality, lexicon, and discourse meaning. *Rivista di linguistica, 15*(2), 277–303.
Moeschler, J. (2011). Causal, inferential and temporal connectives: Why *parce que* is the only causal connective in French. In S. Hancil (Ed.), *Marqueurs discursifs et subjectivité* (pp. 97–114). Rouen: Presses Universitaires de Rouen et du Havre.
Moeschler, J., & Reboul, A. (1994). *Dictionnaire encyclopédique de pragmatique*. Paris: Seuil.
Moeschler, J., Jayez, J., Luscher, J.-M., de Saussure, L., Sthioul, B., & Kozłowska, M. (1998). *Le temps des événements: pragmatique de la référence temporelle*. Paris: Kimé.
Moeschler, J., Grisot, C., & Cartoni, B. (2012). Jusqu'où les temps verbaux sont-ils procéduraux? *Nouveaux Cahiers de linguistique française, 30*, 119–139.
Nicolle, S. (1998). A relevance theory perspective on grammaticalization. *Cognitive Linguistics, 9*, 1–35.
Noveck, I. A., & Sperber, D. (2004). *Experimental pragmatics*. Basingstoke/New York: Palgrave Macmillan.
Partee, B. (1984). Compositionality. *Varieties of Formal Semantics, 3*, 281–311.
Radden, G., & Dirven, R. (2007). *Cognitive English grammar*. Amsterdam: John Benjamins.
Reboul, A. (1992). *Rhétorique et stylistique de la fiction*. Nancy: Presses universitaires de Nancy.
Reboul, A., & Moeschler, J. (1995). Le dialogue n'est pas une catégorie naturelle scientifiquement pertinente. *Cahiers de linguistique française, 17*, 229–248.
Reboul, A., & Moeschler, J. (1996). Faut-il continuer à faire de l'analyse de discours? *Hermes, 16*, 61–92.
Reboul, A., & Moeschler, J. (1998). *Pragmatique du discours. De l'interprétation de l'énoncé à l'interprétation du discours*. Paris: Armand Colin.
Reichenbach, H. (1947). *Symbolic logic*. Berkeley: University of California.
Smith, N. (1990). *Observations on the pragmatics of tense* (UCL Working Papers in Linguistics 2, pp. 113–114).
Smith, C. (2001). Discourse modes: Aspectual entities and tense interpretation. *Cahiers de grammaire, 26*, 183–206.
Smith, C. S. (2003). *Modes of discourse: The local structure of texts*. Cambridge: Cambridge University Press.
Smith, C. S. (2010). The domain of tense. In R. Meier, H. Aristar-Dry, & E. Destruel (Eds.), *Text, time, and context* (pp. 161–182). Dordrecht: Springer.
Sperber, D., & Wilson, D. (1986/1995). *Relevance: Communication and cognition*. Oxford: Blackwell.

Sperber, D., & Wilson, D. (1998). The mapping between the mental and the public lexicon. In P. Carruthers & J. Boucher (Eds.), *Language and thought: Interdisciplinary themes* (pp. 184–200). Cambridge: Cambridge University Press.

Tomasello, M. (2000). First step toward a usage-best theory of language acquisition. *Cognitive Linguistics, 11*, 61–82.

Wagner, R. L., & Pinchon, J. (1962). *Grammaire du français classique et moderne*. Paris: Hachette.

Weinrich, H. (1973). *Le temps: le récit et le commentaire*. Paris: Seuil.

Wilson, D. (2011). The conceptual-procedural distinction: Past, present and future. In V. Escandell-Vidal, M. Leonetti, & A. Ahern (Eds.), *Procedural meaning: Problems and perspectives* (pp. 3–31). Bingley: Emerald Group Publishing Limited.

Wilson, D., & Sperber, D. (1993). Linguistic form and relevance. *Lingua, 90*, 1–25.

Wilson, D., & Sperber, D. (1998). Pragmatics and time. In R. Carston & S. Uchida (Eds.), *Relevance theory: Applications and implications* (pp. 1–22). Amsterdam: John Benjamins.

Subject Realization in Japanese Conversation by Native and Non-native Speakers: Exemplifying a New Paradigm for Learner Corpus Research

Stefan Th. Gries and Allison S. Adelman

Abstract In the field of Learner Corpus Research, Gries and Deshors (Corpora 9(1):109–136, 2014) developed a two-step regression procedure (MuPDAR) to determine how and why choices made by non-native speakers differ from those made by native speakers more comprehensively than traditional learner corpus research allows for. In this chapter, we will extend and test their proposal to determine whether it can also be applied to pragmatic and grammatical phenomena (subject realization/omission in Japanese), and whether it can help study categorical differences between learner and native-speaker choices; we do so by also showing that the more advanced method of mixed-effects modeling can be very fruitfully integrated into the proposed MuPDAR method. The results of our study show that Japanese native speakers' choices of subject realization are affected by discourse-functional factors such as givenness and contrast of referents and that, while learners are able to handle extreme values of givenness and marked cases of contrast, they still struggle (more) with intermediate degrees of givenness and unmarked/non-contrastive referents. We conclude by discussing the role of MuPDAR in Learner Corpus Research in general and its advantages over traditional corpus analysis in that field and error analysis in particular.

Keywords Learner corpora • Regression modeling • Subject realization • Japanese • Givenness and contrast

St.Th. Gries (✉) • A.S. Adelman
Department of Linguistics, University of California, Santa Barbara,
Santa Barbara, CA 93106-3100, USA
e-mail: stgries@linguistics.ucsb.edu

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2,
DOI 10.1007/978-3-319-06007-1_3, © Springer International Publishing Switzerland 2014

1 Introduction

1.1 The State of the Art in Learner Corpus Research

Given the increasing availability of learner corpora, learner corpus research (LCR) is a growing sub-field of corpus linguistics. Much of the work done in LCR is concerned with "bring[ing] out the words, phrases, grammatical items or syntactic structures that are either over- or underused by the learner" (Granger 2002: 132) and/or seeks to "uncover factors of 'foreign-soundingness'" (Granger 1996: 43), specifically "foreign-soundingness *even in the absence of downright errors*" (Granger 2004: 132, our emphasis).

The way much of the work in LCR proceeds can be summarized as follows:

> In either case the learner deviates in plus or minus from a certain statistical norm which characterizes native performance in a particular language. To ascertain such an error [though see below], one has to perform a *quantitative contrastive study of texts* written by native users of a particular language and by a non-native user of the same language and *compare the frequencies of use* of the investigated forms. (Krzeskowski (1990: 206), quoted from Granger 1996: 45, our emphases)

That is to say, one generates concordances of a phenomenon in question, determines the frequencies with which it is attested both in native language (NL)/ native-speaker (NS) data and in non-native speaker (NNS) data, and compares them to determine whether, relative to the NS standard, the NNS over- or underuse the linguistic unit under consideration. Examples include

- Aijmer (2005), who explores the frequencies of use of modal verbs in NS English (in the LOCNESS corpus) and NNS English (in the Swedish component of the ICLE corpus) with multiple chi-squared tests;
- Altenberg (2005), who discusses frequencies/percentages of uses of English *make* and Swedish *göra* in four different constructional patterns and an 'other' category;
- Cosme (2008), who discusses (cross-linguistic) transfer-related issues based on the over-/underuses of adverbial and adnominal present/past participle clauses by French- and Dutch-speaking learners of English;
- Hundt and Vogel (2011), who explore the frequencies of progressives in data from corpora covering English as a NL, English as a second language, and English as a foreign language on the basis of likelihood-ratio tests;
- Hasselgård and Johansson's (2012) case study of the use of *quite* in the LOCNESS corpus and four components of the ICLE Corpus (Norway, Germany, France, and Spain) involving chi-squared tests comparing *quite*'s frequency (both on its own and with a colligation) from the ICLE components to the LOCNESS frequency;
- Neff van Aertselaer and Bunce (2012), who discuss the frequencies of reporting verbs in the Spanish component of the ICLE corpus and a small academic-writing corpus compiled from Spanish EFL students;
- Rogatcheva (2012), who compares the uses of present perfects by Bulgarian and German learners of English in the corresponding parts of ICLE; etc.

While the above kinds of studies appear to be what is currently the state of the art, this state of the art is severely lacking even if compared to two quite basic and very reasonable desiderata stated a long time ago. First, chi-squared tests of goodness-of-fit (of mere frequencies of occurrence) or of independence (of frequencies of co-occurrence) are certainly not the "massive statistical research" called for by Krzeszowski as early as 1990 (p. 212). Second, they are also not "comparing/contrasting what non-native and native speakers of a language do *in a comparable situation*" (Pery-Woodley 1990: 143, quoted from Granger 1996: 43, our emphasis). Both of these problems have a similar root, namely the fact that many studies reduce the context of a phenomenon under investigation to maximally one co-occurring factor/predictor, such as when Altenberg (2005) explores the use of *make* based on one predictor – patterns that *make* co-occurs with – or when Hasselgård and Johansson (2012) explore the use of *quite* based on one predictor – its colligation. However, this is neither comprehensive enough – surely the use of *make* or *quite* is co-determined by more than this one predictor – nor does a single predictor make the situations of use of *make* and *quite* comparable. As Gries and Deshors (2014) argue on the basis of the alternation of *may* vs. *can*,

> for example, the choice of the modal verbs *can* vs. *may* is determined by 15 or so different factors $F_{1–15}$ including syntactic characteristics of the clause and various morphological and semantic features of the subject [...], and maybe also by the circumstances of production, which we may call register. Thus, the traditional interpretation of "in a comparable situation" leads to the somewhat absurd assumption that we compare uses of NS and NNS that are completely different in terms of $F_{1–15}$ and only share the single factor that they were produced in an essay-writing situation in school.

Without wanting to be alarmist or polemic, it is not clear how the study of any phenomenon P that is determined by 15 or so different linguistic $F_{1–15}$ can be studied with over-/underuse counts at all. If a study on P bases a whole theory about how learners' use of X is affected by L1 influence/interference, teaching materials, etc. on just F_1 while completely ignoring $F_{2–15}$, how insightful can it be? Again, Gries and Deshors (2014) is instructive and merits a long-ish quote:

> From this perspective, it is obvious how lacking mere over-/underuse counts are: If a learner used *may* 10 % less often in a corpus file than a native speaker did, that discrepancy may be completely due to individual cases where closer inspection would reveal that, in many of these specific situations, a native speaker would also not have used *may*. Maybe the learners even wrote about the same topic as the native speaker but used more negated clauses than the native speaker. Negation is inversely correlated with the use of *may* so the fact that the learner used *may* 10 % less often than the native speaker says nothing about proficiency regarding *can*/*may* or over-/underuse as it is traditionally used – that 10 % difference is completely due to the learners' use of negations and, crucially, had the native speaker chosen negations as well, he would have exhibited the same perceived dispreference of *may*.

1.2 First Improvements

Given the above severe shortcomings of the state-of-the-art over-/underuse counts, what can been done to address this? So far, three main kinds of suggestions stand

out.[1] One kind is exemplified by Tono (2004) or Collentine and Asención-Delaney (2010). The former studies verb subcategorization patterns by Japanese learners of English and is particularly instructive in how he takes interactions between predictors into consideration.

The latter explore the use of *ser*/*estar* + adjective using multifactorial regression modeling. Their work is highly interesting as it is one of the few published LCR studies that uses a regression-based approach and, thus, cover a large number of linguistic and contextual factors. Unfortunately, Collentine & Asención-Delaney's methodology has critical problems: one conceptual in nature, two statistical. The conceptual problem is that their study involves two regression models – one for *ser* + adjective and a separate one for *estar* + adjective – when what they should have done is one regression model for all the data including a predictor Verb: *ser* vs. *estar* that is allowed to interact with all others. This would have allowed them to see whether any effects differ significantly between the two verbs. As for the statistical problems, a somewhat subjective one is that best-subsets analyses are far from uncontroversial and have been surpassed by other methods (e.g., Lasso and Least Angle regressions). However, the authors do not provide enough information on how their statistical analysis proceeded, but typical implementations of this method neither include interactions between predictors in their computations nor allow for non-linear effects, which is problematic since we know from now two decades of research on lexical and syntactic alternation phenomena in linguistics that they usually involve interactions between predictors and sometimes also non-linear effects.[2]

The second kind of approach addresses several of Collentine & Asención-Delaney's problems and involves regression analyses of corpus data where

- the choice constituting phenomenon *X* to be studied is the dependent variable;
- many linguistic/contextual variables are the independent variables;
- an additional independent variable is the L1 of the speaker, which should minimally compare NS data to one NNL, but multiple NNLs would be better and more in line with, for example, Granger's (1996) Contrastive Interlanguage Analysis;
- the L1 variable can interact with all other predictors because only that will bring out whether any linguistic/contextual variable differs across the L1s.

This approach has been discussed in detail in Gries and Deshors (2014) and Gries and Wulff (2013) as well as several conference papers by the latter two authors.

The most fine-grained approach so far, however, is the so-called MuPDAR approach (Multifactorial Prediction and Deviation Analysis with Regressions) of

[1] We are disregarding here the large body of multifactorial work done by Crossley, Jarvis, and collaborators (cf. in particular Jarvis & Crossley 2012) because much of that work focuses on detecting the L1 of a writer rather than, as here, understanding any one particular lexical or grammatical choice in detail.

[2] An additional problem may involve the fact that the authors used a linear regression on data that might violate the assumptions of such regressions. However, we were unable to infer from the paper what the dependent variable was – possibly a frequency of *ser*/*estar* + adjective per file? – so the above has to remain speculation for now.

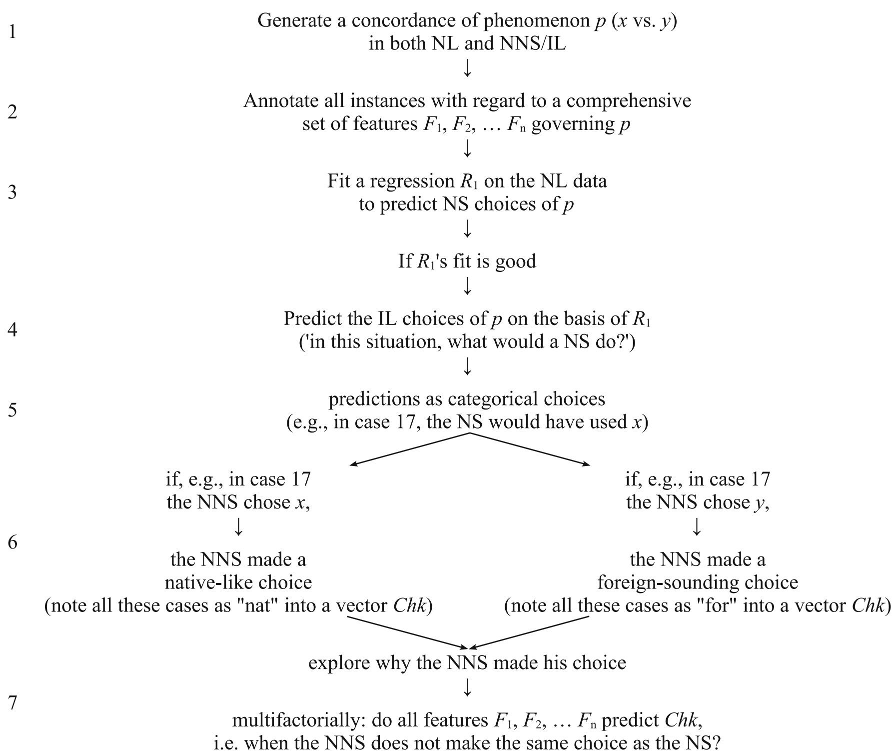


Fig. 1 Flowchart of the MuPDAR approach (Gries and Deshors 2014)

Gries and Deshors (2014). It involves a two-step regression procedure that offers an unprecedented level of precision in the analysis of learner language and is represented in a simplified version in Fig. 1.

First, one generates a concordance of phenomenon *P* and annotates it for an ideally large number of factors/predictors F_{1-n} that influence *P*. Then, P is modeled in a first regression R_1, but only on the basis of the NS data. If that regression model fits the data well, then its regression equation does a good job at quantifying each predictor's importance and predictive power for *P* and that means one can apply it to the NNS data. This is the first most essential step: What it does is answer for every NNS choice with regard to *P* the question "what would a native speaker have done?" These answers, i.e., the predicted NS choices, can then be compared to what the NNS did: either the NNS made the same choice as is predicted from the NS data, then he 'got it right', or the NNS made a choice that differs from what a NS would have done, in which case the NNS choice may not be prescriptively wrong, but at least not native-like. The final step then consists of a second regression R_2, in which one tries to identify which of the factors/predictors F_{1-n} result in the NNS making non-native-like choices. The results of this regression R_2 can then be interpreted in various ways; one of the most natural ways is that predictors that lead to higher NNS error rates can be considered 'difficult' for the learners.

1.3 Goals and Structure of the Present Chapter

Gries and Deshors (2014) exemplify the above approach with regard to modal choice by French learners of English. Their approach and results are quite promising but we want to explore two things they have not done. First, Gries and Deshors (2014) actually adopt a finer level of granularity than shown above: Rather than just considering categorically whether a NNS speaker makes a NS choice or not (cf. step 6), they consider the degree to which the NNS did not make a NNS choice. While the latter is arguably more precise, they do not show that the former also yields useful results. Thus, in this chapter, we will test whether their MuPDAR is also useful if one only explores NNS choices in a binary fashion, i.e., whether they correspond to the predicted NS choices.

Second, in their proof-of-concept chapter, they do not utilize the fullest potential of statistical analysis for their data. Specifically, they analyze the choice of *may* vs. *can* with a binary logistic regression model but, while the results are admittedly very promising, their case might have been stronger if they had analyzed the data with a generalized linear mixed-effects model (GLMEM). These models have become increasingly popular in linguistics over the last few years (cf. Baayen 2008; Jaeger 2008) given their ability to

– handle unbalanced designs, i.e., the type of unequal-cell-frequency problems that are emblematic of corpus-linguistic research;
– handle the fact that the data points entered into a corpus-linguistic analysis are often not independent of each other, since one speaker/writer may contribute multiple data points.

Thus, in this chapter, we will test whether the initial success of their MuPDAR approach can be replicated once more advanced GLMEMs are used. Incidentally, this will also be methodologically interesting on its own because of how GLMEMs work. In order to address the relatedness of data points, GLMEMs can provide (speaker-specific) adjustments to the overall intercept of the regression model, the contrasts between levels of categorical predictors, the slopes of numeric predictors, and interactions of predictors. However, the MuPDAR approach involves applying a model that was fit on data from one set of speakers – the native speakers – to a different set of speakers – the non-native speakers – so our analysis will have to take special steps to take this into consideration.

Finally, while Gries & Deshors studied a lexical choice (*may* vs. *can*), we will explore a pragmatic/grammatical choice – subject realization in conversational Japanese.

Section 2 will discuss our corpus data, their annotation, and their statistical analysis using the extension of the MuPDAR approach with GLMEMs. Section 3 will then turn to the results of the analyses. Specifically, Sect. 3.1 provides the results of the first regression model R_1 on the basis of the NS data; Sect. 3.2 briefly discusses the results of applying R_1 to the NNS data, and Sect. 3.3 is then concerned with the second regression model R2, which explores the non-nativeness of non-native speaker choices. Section 4 concludes.

2 Data and Methods

To further explore the MuPDAR approach, we decided to explore the phenomenon of subject realization in Japanese. Subject arguments are not expressed in all Japanese clauses; in fact they are quite often left unrealized, in what has been discussed as "pro-dropping," "ellipsis," or "zero anaphora" (e.g., Clancy 1980; Hinds 1982); cf. (1) for examples of one clause with a realized subject ((1)a) and one without ((1)b).

(1)	a.	*uchi-no*	*ryoushin-wa,*	
		1SG-GEN	parents-TOP	
		Shizuoka-ni		*sunde-i-te,*
		Shizuoka-LOC		live-PROG-CONJ
		'my parents live in Shizuoka, and . . .'		
	b.	*muzukashi-i*	*to*	*omo-u.*
		difficult-NPST	QUOT	think-NPST
		'(I) think (it)'s difficult.'		

Shibatani (1985: 839) describes "PRO-dropping" as a process in Japanese – and Romance languages – in which "pronouns are omitted [. . .] because of their recoverability from the context." Ono and Thompson (1997: 484) have proposed that predicates should not be seen as having "obligatory" arguments or "slots" calling for either a mentioned referent or a "zero" (although the intended referents may be easily inferred from pragmatic context). Subsequent studies, claiming that unexpressed referents can usually be inferred from context, have therefore argued for the importance of examining this phenomenon only in the discourse contexts of interactional or conversational environments (Takagi 2002); in Sect. 2.1.1 we will discuss the corpus data that we will analyze in the present chapter.

Native speakers' realization of subjects in Japanese is based on many nuanced discourse-pragmatic factors which are likely to be difficult for NNS, particularly those with less experience speaking conversational Japanese. Given how speakers have to navigate information-structural demands and the recoverability and/or inferrability of referents in conversational real time, we assume that NS' patterns of subject realization are influenced by discourse-pragmatic factors such as givenness and contrast; accordingly, in this chapter we will explore if and how choices of subject realization or non-realization differ between NS and NNS speakers of Japanese and how these are affected by, or at least correlated with these two factors; in Sect. 2.1.2 we will therefore discuss our annotation of the corpus data.

2.1 Data

2.1.1 The Corpus Data

Data for this corpus of Japanese NS and NNS conversations was collected in various cities across Japan in the fall of 2011. The corpus consists of four hours of

conversational data, comprising twelve 20-min conversations, each between one NS and one NNS of Japanese. The 12 conversations were carried out by 24 unique subjects, who volunteered to participate in pairs of two; in all cases these pairs were self-described "friends" (eight pairs), "close friends" (three pairs), or spouses (one pair).

In Japanese, speakers' relationships and social status are relevant to the style or register of spoken language used; by selecting only volunteer pairs of friends or spouses, we could ensure the near-consistent use of casual-register Japanese, rather than the distinct polite-register Japanese, throughout the corpus. While many Japanese language textbooks or L2-learning approaches focus primarily on formal or polite registers of the language (typically used among people who have only recently met), communication that takes place in such social settings likely constitutes only a small fraction of the total amount of linguistic interaction in which Japanese native speakers – and many non-native speakers – participate. Previous Japanese L2 speaker corpora have consisted of formal Japanese in artificial interview settings (Hypermedia Corpus of Spoken Japanese; cf. http://www.env.kitakyu-u.ac.jp/corpus/docs/index.html), as well as written Japanese compositions (Learner's Language Corpus of Japanese; cf. http://cblle.tufs.ac.jp/llc/ja/), but no corpus to our knowledge has attempted to capture casual everyday conversation among NNS and NS speakers who are already well-acquainted with each other (e.g., close friends or spouses), in more natural settings.

The 12 NNS participants had as their L1s either English (8), Korean (2), or Chinese (2); the native English speakers were from the U.S., the UK, Canada, and Australia. Most had taken some coursework in Japanese, while some had learned the language primarily through self-study with textbooks and conversations with Japanese friends. All but one of the NNS participants had been studying Japanese for at least 4 years (self-reported study times ranged from 2 to 41 years). All but two of the NNS participants had been living in Japan for at least 3.5 years (self-reported time spent living in Japan ranged from 1 month to 26 years).

Eight of the twelve NS participants came from the central Chuubu and Kansai regions of Japan; two others were from Okayama prefecture, and two were from Tokyo. Most described themselves as speaking regional dialects, with the two from Tokyo reporting that they spoke *hyoujungo*, or standard Japanese based on the Tokyo dialect.

The recordings in the corpus were transcribed in Romanized Japanese in a slightly-adapted version of DT2 (cf. Du Bois 2006) by the second author; each transcription was thoroughly double-checked by a native Japanese speaker.[3] The corpus contains a total of 13,555 intonation units, and a total of 6,873 clauses (55 % verbal predicates; 24 % nominal predicates; 21 % adjectival predicates).

[3] We thank Nobutaka Takara and Mikuni Okamoto for their help in transcribing the corpus data.

2.1.2 The Annotation

We included most of the clauses from the corpus in our sample, excluding those for which a particular subject referent could not be identified, as explained further below. This resulted in a sample of 5,952 sentences. These were then annotated with regard to the following set of variables. First, every clause was coded for the variable Speaker, i.e., a variable indicating whether the speaker of the clause is a native Japanese speaker (*N(J)S*) or a non-native speaker (*NNS*).

Second, every clause was coded for the variable Givenness, i.e., an interval-scaled variable reflecting the givenness of the subject referent on a scale from 0 to 10. High values (*10*, *9*, *8*, etc.) indicate that the referent is highly given (e.g., the referent has been mentioned directly or indirectly (mentioned overtly or referred to implicitly) in the previous clause (*10*), one clause back (*9*), two clauses back (*8*), etc.), while lower values reflect a greater distance to the last mention (e.g., the referent has been mentioned 9 clauses back (*2*), the referent has been mentioned 10 clauses back (*1*), and a value of *0* indicates that the referent has not been mentioned at all in the 10 preceding clauses). However, given the nature of the data – conversations of two speakers – the referents of first- and second-person expressions were always coded with a *10*.

Third, every example was annotated for Contrast, i.e., a variable representing whether the subject is contrastive (*yes*) or not (*no*). The annotation of Contrast required a detailed inspection of the clauses' contexts. For example, whereas some *wa*-marked NPs act as topics, two *wa*-marked NPs in two clauses in a row leads each of those clauses to have a contrastive structure (Iwasaki 2002: 244), as exemplified in (2), where both clauses are coded as having contrastive subjects.

(2)	*de*	*hitori-wa*	*tabete-i-mashi-ta.*
	and	one.person-TOP	eat-PROG-POL-PST
	'so one person person was eating.'		
	hitori-wa	*matte-i-mashi-ta.*	
	one.person-TOP	wait-PROG-POL-PST	
	'another(/one) person was waiting.'		

We did not code for "propositional contrast," meaning instances where the entire clause is contrasted with another proposition, rather than one particular element in the clause being marked as contrastive (Kuno 1973: 46–47). We coded only for contrastive subject/topic arguments (rather than contrastive object arguments or propositional contrast), i.e., only for when two or more subjects/topics were being contrasted with each other, usually with respect to the same predicate. For example, both of the following clauses were coded as contrastive because of the affirmative/negative polarity contrast of only one particular element in each clause against the other (this is not an example of propositional contrast because both clauses have the same predicate).

(3)

nanka	*kekkou*	*shaber-u*	*ko*	*mo*	*i-tari,*
DM	quite.a.bit	speak-NPST	kid	too	exist-REP

'like there are students who speak quite a bit, and,'

shaber-e-nai	*ko*	*mo*	*i-tari*	*shite,*
speak-POT-NEG	kid	too	exist-REP	light.verb

'and there are also students who can't speak.'

In addition, arguments were marked as contrastive when they involved contrastive topics, when they were subjects of the inherently contrastive construction *(no) hou ga*, or when they involved the *yori* 'compared to' construction.

For some predicates whose subjects are not realized, it is impossible to identify a particular referent as the intended subject; this is sometimes – though not always – due to predicates being part of "fixed expressions with different degrees of lexicalization" (Ono and Thompson 1997: 485). For any predicates for which we could not identify a particular referent as the subject, we labeled those clauses as "uncodeable" and did not include them in our sample.

Finally, each clause was coded for the variable SubjReal, to reflect whether a subject was realized (*yes*) or not (*no*).

In addition to these fixed-effects predictors, we also included annotation for a random effect representing the identity of the speakers (SpeakerID) so that the fact that the data points are not independent but may involve speaker-specific effects is taken into consideration. The corpus consists of 12 recordings, each between a native and a non-native speaker; thus there are 24 individual speaker IDs.

2.2 Statistical Evaluation

In this section, we outline the statistical evaluation of the above-mentioned corpus data. We proceed in three steps: the description of the model fitting/selection process of R_1 using the NS data, its application to the NNS, and the model fitting/selection process of R_2.

2.2.1 Regression R_1: Exploring the Choices Made by NS

Our model fitting/selection process follows the logic outlines by Zuur et al. (2009: Ch. 5).[4] That is, we first determine the random-effects structure of the model, then the fixed-effects structure. As for the former, we begin with a maximal model that

[4] By virtue of the complexity of the statistical methods involved, this section can only be rather technical in nature, plus space constraints do not permit exhaustive definitions and discussion of all the statistical technical terms. We therefore refer the reader to Baayen (2008: Ch. 7), Crawley (2013: Ch. 9, 19), Faraway (2006: Ch. 8–10), and Zuur et al. (2009: Ch. 5).

was fit to the 3263 NS data points only (using REML estimates and the function lmer from the R package lme4 (version 0.999999-2); cf. Bates et al. 2013) and included

- SubjReal: *no* vs. *yes* as the dependent variable;
- Givenness: *0*, *1*, ..., *9*, *10* and Contrast *no* vs. *yes* and their interaction as fixed-effects predictors, where, to allow for possible curvature in the effect of Givenness, the maximal model included Givenness as a polynomial to the third degree;
- random intercepts and slopes for all predictors and their interactions as random effects.

Using likelihood-ratio tests, the random-effects structure of this model is reduced to the minimal adequate one, i.e., the one that did not allow further simplification. After that, we proceed with an analogous reduction of the complexity of the fixed-effects structure using likelihood-ratio tests of ML fits to the final minimal adequate model. The quality of this model is then assessed by means of an overall likelihood-ratio chi-squared significance test, the model's classification accuracy, and its *C*-score; the nature of the effects of this final model is interpreted with plots of predicted probabilities of subject realization both separately for each speaker and as an overall trend.[5]

2.2.2 Applying R_1 to the NNS Data

The next step involves applying the regression model R_1 to the NNS data. Crucially, R_1 involves speaker-specific effects, but since the NNS data stem from different speakers, we only use the fixed-effects part of R_1 to answer the following question for every NNS data point: "would a native speaker have realized the subject here, yes or no?" The fit of the NS model to the NNS data is also quantified with a classification accuracy and a *C*-score.

2.2.3 Regression R_2: Exploring the Choices Made by NNS

Given the results from Sect. 2.2.2, we can determine for each of the 2689 NNS data points whether the NNS chose what was predicted as the most likely NS choice. The results of this comparison are represented in a variable called Correct: *no* (the NNS made the predicted NS choice) vs. *yes* (the NNS did not make the predicted NS choice). The variable Correct then is the dependent variable in the second

[5] Strictly speaking, if one does a MuPDAR analysis in which R1 is really only used for prediction, then one does not really have to apply Occam's razor rigorously to eliminate non-significant/collinear predictors that much because, within MuPDAR, the point of R1 is not to actually interpret R1's coefficients.

Table 1 Results of R_1 (predicted level of SubjReal: *yes*)

Fixed effects				
Predictor	Estimate/coefficient	Std. error	z	p_{deletion}
Intercept	1.25353	0.17581	7.130	<<0.0001
Givenness	−0.33058	0.02043	−16.180	<<0.0001
Contrast (*no* → *yes*)	0.07159	0.36953	0.194	0.846
Givenness * Contrast (*no* → *yes*)	0.34734	0.05130	6.771	<<0.0001
Random effects				
Adjustment to overall intercept (Speaker)	$sd = 0.477134$			
Adjustment to slope of Givenness	$sd = 0.054543$			

regression model fitting/selection process R_2, which proceeds as before: we include the same fixed-effects predictors and random effects as for R_1, first determine the minimal adequate random-effects structure (with likelihood-ratio tests of REML fits), and then the minimal adequate fixed-effects structure (with likelihood-ratio tests of ML fits). Finally, we compute the final model's significance test and classification accuracy and visualize its results in terms of the predicted probabilities of the NNS making the choice that the NS would have made.

3 Results

In this section, we present the multitude of results of the statistical analyses; we proceed analogously to Sect. 2.2.

3.1 *The Results of* R$_1$*, the Regression on the NS*

The results of the first regression, R_1, applied to the NS data only, indicate a good fit. The minimal adequate model we arrived at after the model selection process reflects a highly significant correlation between its predictors and the NS choices of subject realizations: likelihood-ratio chi-squared = 192.13, $df = 3$, $p < 0.0001$. Table 1 represents the results for the fixed and random effects in the model.

The results in Table 1 already indicate that

- if Givenness increases, the probability of a subject being realized decreases (note the negative sign of the coefficient of Givenness);
- Contrast on its own has no effect on subject realization;
- the interaction of Givenness and Contrast is highly significant and in fact annuls the effect of Givenness in isolation when Contrast is *yes*

As usual, however, these effects are much easier to comprehend from a visual representation such as Fig. 2. In both panels of Fig. 2, Givenness is represented on

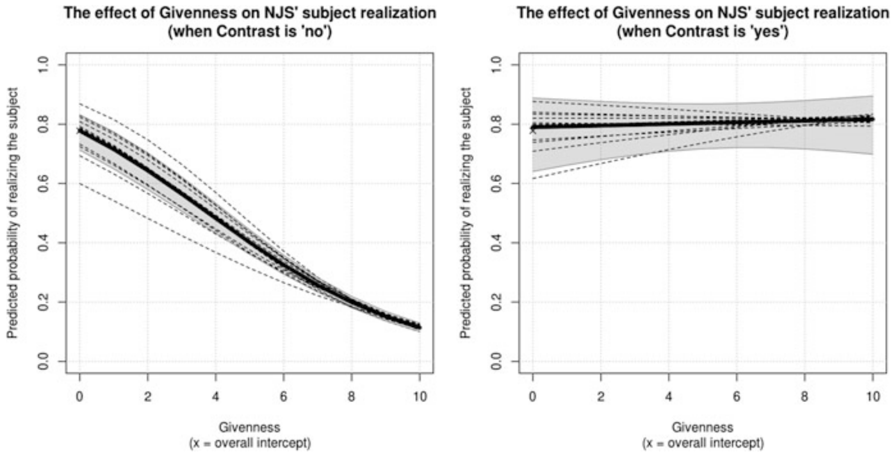


Fig. 2 The effect of the interaction Givenness * Contrast on the predicted probability of NJS' realizing the subject

Table 2 Classification accuracy of R_1 when applied to the NS

	Pred. SubjReal: no	Pred. SubjReal: yes	Totals
Obs. SubjReal: no	2,279	148	2,427
Obs. SubjReal: yes	345	491	836
Totals	2,624	639	3,263

the x-axis while the predicted probability of a subject being realized is represented on the y-axis; to provide a fine-grained resolution of the results, we indicate both the results for every speaker individually (with dashed grey lines) and the results for all speakers (a heavy black line with its grey confidence interval). The left panel shows the effect of Givenness when Contrast is *no*, and there is a strong and clear trend such that, the more given the referent of the subject, the less likely it will be expressed overtly; the speaker-specific results show that this effect holds for all speakers (but of course to varying degrees). The right panel shows the effect of Givenness when Contrast is *yes*; the essentially flat regression line indicates that Givenness has no effect on subject realization when Contrast is *yes* – whatever the value of Givenness, in contrastive settings subjects are very likely to be realized. In this panel, we do find some subject-specific variation: some slopes exhibit an upward trend, some a downward trend, but since a random effect Givenness: Contrast|Speaker did not reach standard levels of significance, the overall conclusion – Givenness has no effect on subject realization when Contrast is *yes* – still stands (Fig. 2).

Even though the final model contains only one significant highest-level predictor, the classification accuracy of the model amounted to 84.9 %, which is highly significantly better than the chance-level baseline of 61.9 % ($p_{\text{binomial test}} < 10^{-100}$); consider Table 2 for the classification matrix resulting from the predictions of R_1. The more precise C-value for this model is 0.82, thus exceeding Harrell's (2001: 248) threshold of 0.8 for good models.

Table 3 Classification accuracy of R_1 when applied to the NNS

	Pred. SubjReal: no	Pred. SubjReal: yes	Totals
Obs. SubjReal: no	1,767	124	1,891
Obs. SubjReal: yes	305	493	798
Totals	2,072	617	2,689

Table 4 Results of R_2 (predicted level of Correct: *yes*)

Predictor	Estimate/coefficient	Std. error	z	$p_{deletion}$
Fixed effects				
Intercept	1.76	0.1423	12.369	<<0.0001
Givenness	11.4229	2.6543	4.304	<<0.0001
poly(Givenness, 2)	10.8993	2.3960	4.549	<<0.0001
Contrast (*no* → *yes*)	0.4872	0.2552	1.909	0.04573
Random effects				
Adjustment to overall intercept (Speaker)	$sd = 0.44668$			

3.2 The Results of Applying R_1 to the NNS Data

Given the good fit of R_1 to the NS data, we proceeded by generating predictions of subject realizations for the NNS data. Crucially and as mentioned above in Sect. 2.2.2, the predictions for the NNS were based only on the fixed effects listed in Table 1, i.e., the speaker-specific random effects of R_1 were not included given that the NNS are different speakers. Nevertheless, R_1 was able to predict the subject realizations of the NNS nearly exactly as well as those of the NS; consider Table 3 for the classification matrix; the accuracy of the model is 84 %, which is highly significantly better than the chance-level baseline of 58.3 % ($p_{\text{binomial test}} < 10^{-100}$), and $C = 0.8$.

3.3 The Results of R_2, the Regression on the NNS

The results from Table 3 then lead to the final step, the regression R_2 that was fit to predict when the NNS would make a choice differing from that predicted from the NS data; that is, the dependent variable here was Correct: *no* (1,767 + 493-2,260 cases) vs. *yes* (124 + 305 = 429 cases). The minimal adequate model was again highly significant: likelihood-ratio chi-squared = 37.18, $df = 3$, $p < 0.0001$. The results for all fixed and random effects are represented in Table 4; interestingly, the effect of Givenness is not best represented with a straight line but rather with a curved line resulting from a polynomial to the second degree.

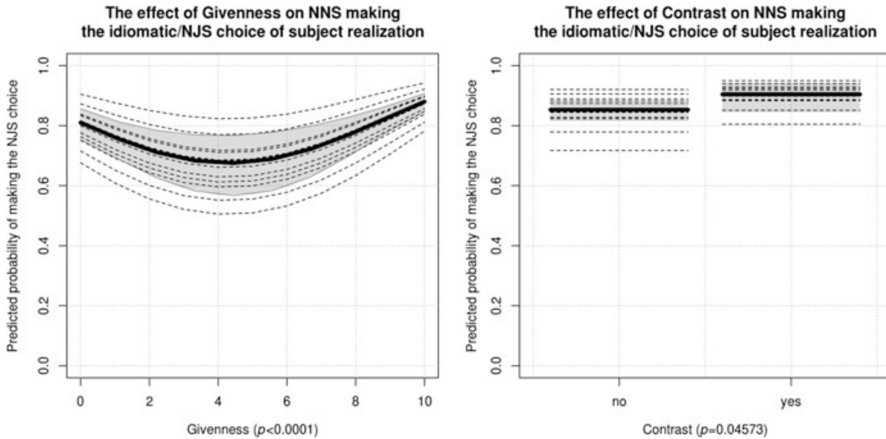


Fig. 3 The significant main effects of Givenness and Contrast on whether NNS make the same subject realization choice a NJS is predicted to have made

Given the curved nature of the effect of Givenness and its being fit with orthogonal polynomials, it is necessary to visualize the results as in Fig. 3. In both panels, the predicted probability of the NNS making the same choice that the NS would have made is represented on the *y*-axis, and again we provide each speaker's prediction with dashed lines and the overall estimate with a heavy line and a grey confidence interval. In the left panel, the predictor Givenness is on the *x*-axis and the result shows that the NNS are most likely to make the NS choice with extreme values of Givenness: i.e., when the referent of the subject is completely new or completely given (in the sense of 'having been mentioned just before'). On the other hand, when the referent of the subject is intermediately given, then the NNS are more likely to not make the subject realization choices a NS would have made. In a nutshell, the NNS can handle the extreme cases, but not (yet) the middle ground.

As for the effect of Contrast, it is relatively weak and only just about significant, but again its results make sense: In the more marked communicative situation with a contrastive subject referent, which can be considered 'more extreme' than the unmarked case, the NNS make choices that are more in line with what NS would have done. On the other hand, when the referent of the subject is not marked (in the sense of 'not being contrastive'), the NNS struggle more with making NS choices.

4 Discussion and Concluding Remarks

4.1 Interim Summary and Implications of the Analysis

The results of the present analysis are strong evidence for the feasibility of the LCR method proposed here, the MuPDAR approach. All regression results are at least

significant and come with high degrees of predictive power/classification accuracy. R_1 shows that NS choices to realize the subject are strongly influenced by Givenness and its interaction with Contrast in ways that are compatible with previous findings regarding discourse givenness or inferrability in a wide variety of languages – given or highly inferrable referents are often not lexically realized – and with what can reasonably be expected for Contrast – referents that are to be highlighted contrastively are realized no matter their givenness. At the same time, R_2 shows that the learners in this study have been able to extrapolate these NS preferences, but not fully yet. Extreme values of givenness/inferrability pose few problems to the learners, as do the distinctions of Contrast: the NNS speakers know what to do with subjects when their referents are completely given, completely new, and contrastive – they still struggle with intermediate degrees of givenness/ inferrability, which not only makes sense since this is an 'uncomfortably grey middle area' on the givenness continuum but also because this kind of scenario happens least often. For discourse cohesion reasons, referents are usually introduced but then also used immediately afterwards, which would result in high values of givenness. But when that does not happen and a referent has been introduced but then left in limbo for 4–6 clauses, then the NNS have problems. The MuPDAR approach has revealed this quite clearly and we submit it is hard to imagine how traditional LCR would have found this (so clearly; cf. below). Follow-up analyses could now also explore the random-effect structure to determine, for example, whether the random intercepts/slopes correlate with relevant characteristics of the speakers, such as their L1s. We did this for the present data but, unlike in Miglio et al. (2013), no correlations between random effects and the speakers were found.

4.2 Where to Go from Here

We hope to have shown that the MuPDAR approach is a powerful and flexible tool for LCR. This second proof-of-concept study shows that (i) MuPDAR cannot only be used with traditional regression modeling but is also naturally extended to advanced mixed-effects modeling and that (ii) MuPDAR yields precise and meaningful results regardless of the resolution of R_2 – categorical deviations of NNS choices from NS choices as in this study or differences in degree as in Gries and Deshors (2014). That being said, there are several obvious next steps. One is that we clearly need more applications of this approach; in an ideal world, this would mean that traditional work in LCR would be re-analyzed to determine its validity.

Second, the method may be further refined. Dan Lassiter (p.c.) suggested considering not only the (categorical or numeric) differences between NNS and NS choices, but also the differences between predicted probabilities of NNS and NS choices, which would make this method relate more seamlessly to variationist sociolinguistic studies. While we have no particular hypotheses about how this perspective would play out, it is certainly worth exploring in future work. In addition, various ways of making the analytical results more robust – cross-

validation with bootstrapping approaches are one possibility – should be explored in due course.

Third, it is also worth pointing out that both existing MuPDAR studies involved a final, minimal adequate regression model R_1 (from which insignificant predictors were trimmed following Occam's razor). The reason for this is that the results of R_1 are then also useful in their own right and can be interpreted linguistically/theoretically. However, if R_1 is really only used for prediction then it would theoretically be possible to not trim the maximal R_1 model and make full use of the fact that its classification accuracy will be slightly higher than that of the minimal adequate model we used here.

At this point, it is instructive to briefly discuss the relation of MuPDAR to error analysis. We believe that the present approach is at least a complement of, if not also a massive improvement over, traditional kinds of error analysis. For instance, some studies – Rogatcheva (2012) is a case in point – explore over- and underuses by having linguists/native speakers perform error-tagging on learner data. This is generally a useful approach given how it allows for, technically speaking, *true positives* (present perfects by NNS where they should be), *false positives* (present perfects by NNS where they should not be), *true negatives* (no present perfects by NNS where they should not be), and *false negatives* (no present perfects by NNS where they should be), and on the basis of such data, one can then compute statistics such as *SOC* (suppliance-in-obligatory-contexts) and *TLU* (target-like-use). On the one hand, this approach is undoubtedly more comprehensive than many previous LCR studies that do not include any context in their counts or that cross-tabulate just a single contextual feature in that the error coders will take more context into consideration in their coding decisions.

On the other hand, the process also suffers from some problems, which have to do with the distinctions that the coders/raters will make. A first problem that may arise is concerned with rater reliability. It has been known for many years now that judgment tasks like these are not only affected by a huge variety of factors (cf. Schütze 1996 for the most authoritative overview showing that) but can also be affected by the stimuli themselves over very short periods of time. For instance, Gries and Wulff (2009) discuss a weak but marginally significant within-subject priming effect that appears to indicate how subjects' preferences for sentence completion change over the course of just a short experiment (even when all other significant predictors are still considered). Similarly, Doğruöz and Gries (2012) find that, over the course of only eight acceptability judgments, subjects became more comfortable with unconventional morphological and lexical patterns. Thus, it is likely that raters' judgments/predictions will be affected as they go over and code many learner choices; at the very least, it is possible that they will and the degree to which they will is unknown. The precision of the MuPDAR approach, by contrast, is not affected by learning, habituation, or fatigue, and given the way that, in this chapter, we used mixed-effects modeling, it even accounts for speaker-specific effects that raters will most likely not be able to attend to.

The above is not to downplay the potential of error analysis, especially not if multiple coders are involved, coding protocols are rigorous, order effects etc. are

Table 5 Observed frequencies of SubjReal ~ Speaker * Contrast

Speaker	Contrast	SubjReal: *no*	SubjReal: *yes*	Totals
NJS	*no*	2,404	741	3,145
	yes	23	95	118
NNS	*no*	1,871	671	2,542
	yes	20	127	147
Totals		4,318	1,634	5,952

controlled, and careful interrater reliability statistics are computed. Nevertheless, even if all of these issues were addressed, MuPDAR still has advantages to offer. For instance, an additional problem of error-coding types of analyses is that most coders will not make as fine-grained distinctions/predictions as the regression because their judgments will at best be binary or categorical predictions about what will or should be used. On the other hand, when R_1 is applied to the NNS data, the MuPDAR approach makes very fine-grained predictions on a continuous probability scale, and when R_2 is computed on the basis of the deviations of NNS choices from the NS predictions as in Gries and Deshors (2014), then this regression, too, operates on a continuous scale. Thus, MuPDAR offers more a precise analysis of the data.

Finally, the error analysis and the resulting identification of, say, false positive and false negatives, in and of itself brings one no closer to an explanation of why the NNS did what they did. In the terminology of the present chapter, what the error-analysis approach does is 'computing R_1 on the NS data and applying it to the NNS data.' However, one then still needs to do R_2 to understand what it is that is responsible for the NNS making choices that are slightly or very much less idiomatic than those of the NS and on that topic, for example, Rogatcheva (2012) does very little. There are undoubtedly many different factors that jointly determine whether or not NS use the present perfect, but her chapter, while (laudably) computing *SOC* and *TLU*, does nothing to shed light on how many such factors there are, what they are, how strongly they affect speaker choices, and what their interactions might be.

Applying MuPDAR to native and learner corpus data is undoubtedly a complex and technical process, which may seem insurmountable to some and off-putting to even more. However, LCR scholars on the whole seem to agree that the corpus-based analysis of NNS language is, if anything, *more* complex than the analysis of NS language, which we already know from decades of alternation research to involve highly complex interactions of factors in multifactorial models. It is therefore utterly illogical to assume that the more complex set of questions regarding NNS language can be tackled with simple over-/underuse frequencies and pairwise chi-squared/log-likelihood ratio tests – complex data sets need techniques that can handle complex data, not methods that reduce the complexity to a level that has nothing to do anymore with what is really happening in the corpus. As a thought experiment, consider the fact that the currently most frequent type of over-/underuse kind of analysis of our data – recall the many studies cited as using such an approach in Sect. 1 – would reduce the analysis of everything that we found in our data to Table 5, presumably coupled with two chi-squared tests

which, strictly speaking, one is in fact not even allowed to compute given that nearly all learner corpus studies are based on data points that are *not* independent, as the chi-squared test would require (which is why we pursued the GLMEM approach).

Against this background, we think it is high time that researchers in LCR begin to embrace tools that do more justice to the complexity they (correctly) claim their data come with. MuPDAR is but one approach to that end, but we believe we have demonstrated it is a powerful one and we hope that it will stimulate many applications exploring the intricacies of NNS language.

References

Aijmer, K. (2005). Modality in advanced Swedish learners' written interlanguage. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition, and foreign language teaching* (pp. 55–76). Amsterdam/Philadelphia: John Benjamins.

Altenberg, B. (2005). Using bilingual corpus evidence in learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition, and foreign language teaching* (pp. 37–54). Amsterdam/Philadelphia: John Benjamins.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Bates, D., Maechler, M., & Bolker, B. (2013). *lme4. Linear mixed-effects models using S4 classes*. http://lme4.r-forge.r-project.org/

Clancy, P. M. (1980). Referential choice in English and Japanese narrative discourse. In W. Chafe (Ed.), *The pear stories: Cognitive and linguistics aspects of narrative production* (pp. 127–202). Norwood: Ablex.

Collentine, J., & Asención-Delaney, Y. (2010). A corpus-based analysis of the discourse functions of *ser*/*estar* + adjective in three levels of Spanish as FL learners. *Language Learning, 60*(2), 409–445.

Cosme, C. (2008). Participle clauses in learner English: The role of transfer. In G. Gilquin, S. Papp, & M. B. Díez-Bedmar (Eds.), *Linking up contrastive and learner corpus research* (pp. 177–200). Amsterdam/Atlanta: Rodopi.

Crawley, M. J. (2013). *The R book* (2nd ed.). Chichester: Wiley.

Doğruöz, A. S., & Gries, S. T. (2012). Spread of on-going changes in an immigrant language: Turkish in the Netherlands. *Review of Cognitive Linguistics, 10*(2), 401–426.

Du Bois, J. W. (2006). *Representing discourse*. Ms., University of California, Santa Barbara.

Faraway, J. J. (2006). *Extending the linear model with R: Generalized linear, mixed-effects and non-parametric regression models*. Boca Raton: Chapman & Hall/CRC.

Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in contrast. Text-based cross-linguistic studies* (pp. 37–51). Lund: Lund University Press.

Granger, S. (2002). A bird's eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3–33). Amsterdam/Philadelphia: John Benjamins.

Granger, S. (2004). Computer learner corpus research: Current status and future prospects. In U. Connor & T. Upton (Eds.), *Applied corpus linguistics: A multidimensional perspective* (pp. 123–145). Amsterdam: Rodopi.

Gries, S. T., & Deshors, S. C. (2014). Using regressions to explore deviations between corpus data and a standard/target: Two suggestions. *Corpora, 9*(1), 109–136.

Gries, S. T., & Wulff, S. (2009). Psycholinguistic and corpus linguistic evidence for L2 constructions. *Annual Review of Cognitive Linguistics, 7*, 163–186.
Gries, S. T., & Wulff, S. (2013). The genitive alternation in Chinese and German ESL learners: Towards a multifactorial notion of context in learner corpus research. *International Journal of Corpus Linguistics, 18*(3), 327–356.
Harrell, F. E., Jr. (2001). *Regression modeling strategies with applications to linear models, logistic regression, and survival analysis*. Berlin/New York: Springer.
Hasselgård, H., & Johansson, S. (2012). Learner corpora and contrastive interlanguage analysis. In F. Meunier, S. De Cock, G. Gilquin, & M. Paquot (Eds.), *A taste for corpora: In honour of Sylviane Granger* (pp. 33–61). Amsterdam/Philadelphia: John Benjamins.
Hinds, J. (1982). *Ellipsis in Japanese*. Edmonton: Linguistic Research, Inc.
Hundt, M., & Vogel, K. (2011). Overuse of the progressive in ESL and learner Englishes – Fact or fiction? In J. Mukherjee & M. Hundt (Eds.), *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap* (pp. 145–165). Amsterdam/Philadelphia: John Benjamins.
Hypermedia Corpus of Spoken Japanese. (2010). http://www.env.kitakyu-u.ac.jp/corpus/docs/index.html. Accessed Fall, 2010.
Iwasaki, S. (2002). *Japanese*. Amsterdam/Philadelphia: John Benjamins.
Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446.
Jarvis, S., & Crossley, S. A. (Eds.). (2012). *Approaching language transfer through text classification explorations in the detection-based approach*. Bristol: Multilingual Matters.
Krzeskowski, T. (1990). *Contrasting languages: the scope of contrastive linguistics*. Berlin & New York: Mouton de Gruyter.
Kuno, S. (1973). *The structure of the Japanese language*. Cambridge, MA: MIT Press.
Learner's Language Corpus of Japanese. (2013). http://cblle.tufs.ac.jp/llc/ja/. Accessed Spring, 2013.
Miglio, V. G., Gries, S. T., Harris, M. J., Wheeler, E. M., & Santana-Paixão, R. (2013). *Spanish lo (s)-le(s) clitic alternations in psych verbs: A multifactorial corpus-based analysis*. Somerville: Cascadilla Press.
Neff van Aertselaer, J. A., & Bunce, C. (2012). The use of small corpora for tracing the development of academic literacies. In F. Meunier, S. De Cock, G. Gilquin, & M. Paquot (Eds.), *A taste for corpora: In honour of Sylviane Granger* (pp. 63–83). Amsterdam/Philadelphia: John Benjamins.
Ono, T., & Thompson, S. A. (1997). Deconstructing 'Zero Anaphora' in Japanese. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society, 23*, 481–491.
Pery-Woodley, M.-P. (1990). Contrasting discourses: Contrastive analysis and a discourse approach to writing. *Language Teaching, 23*(3), 143–151.
Rogatcheva, S. (2012). Perfect problems: A corpus-based comparison of the perfect in Bulgarian and German EFL writing. In S. Hoffmann, P. Rayson, & G. Leech (Eds.), *English corpus linguistics: Looking back, moving forward* (pp. 149–163). Amsterdam: Rodopi.
Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: The University of Chicago Press.
Shibatani, M. (1985). Passives and related constructions: A prototype analysis. *Language, 61*(4), 821–848.
Takagi, T. (2002). Contextual resources for interferring unexpressed referents in Japanese conversations. *Pragmatics, 12*(2), 153–182.
Tono, Y. (2004). Multiple comparisons of IL, L1 and TL corpora: The case of L2 acquisition of verb subcategorization patterns by Japanese learners of English. In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Corpora and language learners* (pp. 45–66). Amsterdam/Philadelphia: John Benjamins.
Zuur, A. F., Ieno, E. N., Walker, N., & Saveliev, A. A. (2009). *Mixed effects models and extensions in ecology with R*. Berlin/New York: Springer.

Jesus! vs. *Christ!* in Australian English: Semantics, Secondary Interjections and Corpus Analysis

Cliff Goddard

Abstract Using corpus-assisted semantic analysis, conducted in the NSM framework (Wierzbicka, Semantics: primes and universals. OUP, New York, 1996a; Goddard, Semantic analysis: a practical introduction, 2nd rev edn. OUP Oxford, 2011), this chapter explores the meanings and uses of two closely-related secondary interjections, namely, *Jesus!* and *Christ!*, in Australian English. The interjections *Shit!* and *Fuck!* are touched on briefly. From a methodological point of view, the chapter can be read as a study in how corpus techniques and semantic analysis can work in tandem; in particular, how interaction with a corpus can be used to develop, refine and test fine-grained semantic hypotheses. From a content point of view, this study seeks to demonstrate two key propositions: first, that it is possible to identify semantic invariants, i.e. stable meanings, even for highly context-bound items such as interjections; second, that it is possible to capture and model speakers' awareness of the degree and nature of the "offensiveness" of secondary interjections, in a Metalexical Awareness component that attaches, so to speak, to particular words. Both these propositions challenge conventional assumptions about the nature and interfacing between semantics and pragmatics. A final question raised in the study is how linguists can come to terms with the fact that people use interjections not only orally but also mentally, in "inner speech".

Keywords Australian English • Corpus analysis • Interjections • Metalexical awareness • Metapragmatics • NSM • Semantic templates • Swearing

C. Goddard (✉)
School of Languages and Linguistics, Griffith University, Brisbane, QLD 4111, Australia
e-mail: c.goddard@griffith.edu.au

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2,
DOI 10.1007/978-3-319-06007-1_4, © Springer International Publishing Switzerland 2014

1 Setting the Scene

1.1 Jesus! *vs*. Christ!

According to several dictionaries (see entries in Table 1), the English interjections *Jesus!* and *Christ!* are identical or near-identical in meaning and usage. Using corpus-assisted semantic analysis, this study will show that such claims are incorrect and there are in fact significant differences in the meanings of these two interjections, and, consequently, significant differences in their usage. In focusing on two small words, and, furthermore, on words which, as interjections, have very limited combinatorial properties, this is a study in micro semantics. At the same time, the study fits into several larger theoretical and methodological agendas, with implications for questions such as: To what extent do highly context-bound items, such as interjections, have specifiable meanings? How can speakers' (ethno) metapragmatic awareness be captured and modelled? How can corpora be used to inform disciplined qualitative analysis (specifically, semantic analysis)?

As interjections *Jesus!* and *Christ!* are sometimes seen as "swear words" and/or "curse words". Along with expressions such as *My God!*, *Goddamn!*, *Damn!*, *Hell!* and others, they are the surviving residue in English of what was once a domain of many powerful, taboo, religious expressions. [Note[1]] These days words from the religious domain have been overtaken in strength, potency and impact by words

Table 1 Dictionary entries for *Christ!* and *Jesus!*, used as interjections

Australian Oxford English Dictionary 2004 (AustOED)	**Christ**: [. . .] ~ ! (vulg.) excl. expr. surprise, impatience, etc. **Jesus** [. . .] ~ (Christ)! (vulg.) excl. expr. surprise, impatience, etc.
Longman Dictionary of Contemporary English 2003 (LDOCE)	**Christ** also **Jesus Christ**, Jesus– *inter sl* (used for expressing annoyance, unwelcome surprise, etc.): *Christ! I've forgotten the keys!* –see JESUS (USAGE) **Jesus** *inter sl* (a strong word used to express surprise, anger, etc.) ■ USAGE Some people, especially those who believe in the Christian religion, are offended by the use of **Jesus** and **Christ** as interjections. **God** is more commonly used and is not felt to be so strong, but some people do not like this use either
Merriam-Webster Online (MWO)	There is no entry for either word as an interjection, but under *Jesus Christ* noun, the following appears as a usage note: ◊ *Jesus Christ*, *Christ*, and *Jesus* are commonly used as interjections to express surprise and anger. These uses are avoided in polite speech

[1] The historical priority of the religious domain has a lexical echo in the fact that the speech-act verbs *swear* and AmEng *curse*/*cuss* (in their "bad language" senses) have descended from speech acts that involved invoking God (cf. Wierzbicka 1987). Incidentally, it can be noted that the relevant meanings of *swear* and AmEng *curse*/*cuss* are not identical but represent somewhat different (i.e. differently construed) speech acts (Goddard Forthcoming).

from the sexual and scatological domains, such as *Fuck!* and *Shit!* (and their numerous variants). The shift away from religious swearing in English has been documented by a number of authors on the social history of swearing and other taboo language, such as Hughes (1998), Allan and Burridge (1991, 2006), Ljung (2011) and Mohr (2013).

The point has often been made that the impact and social functions of "swear words" or "curse words" depends on the assumption that uttering the words in question violates a social convention of some sort; that using these words is, or could be, offensive to some people—if not to one's interlocutor, then at least to some imagined social Other. The dictionary entries given in Table 1 all minimally include a style note such as the OED's 'vulg' [vulgar] or LDOCE's *sl* [slang], and sometimes give a more elaborate comment on usage, such the LDOCE's usage note, which explicitly links the potential offensiveness of *Jesus!* and *Christ!* to belief in the Christian religion. Those with knowledge of the Christian tradition will know that taking God's name "in vain" is prohibited under the Third Commandment, and some may know that the making of oaths (typically with invocations to God) has been a matter of life or death at different times in Christian history. Nevertheless, studies of "degrees of offensiveness" show that for most people today religious swearing/cursing is only mildly or minimally offensive. From an ethnopragmatic point of view (Goddard 2006, in press), the question arises: How do ordinary English speakers think about potential lexical offensiveness? How can metapragmatic knowledge of this kind be modelled in ways that are intuitively natural for speakers?

"Swearing" and "cursing" in general have been the subject of various studies in sociolinguistics, pragmatics and im/politeness studies, psycholinguistics and neurolinguistics (Jay 1992, 2000; Jay and Janschewitz 2008; Van Lancker and Cummings 1999; Holmes and Stubbe 2003; McEnery 2006; Norrick 2009; Stapleton 2010; Culpeper 2011; Beers Fägersten 2012). Some of these studies, notably McEnery (2006) and Beers Fägersten (2012), have involved heavy reliance on spoken corpora or observational data of naturally-occurring speech. An important general finding is that most swearing is apparently not confrontational, rude or aggressive, but rather falls under the broad heading of social or conversational swearing. The social functions of using swear words, such as constructing and demarcating social and gender identities, and enacting mock impoliteness and humour, are increasingly attracting the attention of researchers.

What is conspicuously lacking in this literature is any systematic attention to semantics. The general assumption seems to be that "swearing" expressions do not have any discernable semantic content but function exclusively in the pragmatic dimension. The main exception to this generalisation is to be found in the work of NSM linguists (e.g. Wierzbicka 1992, 1996b, 2002; Kidman 1993; Goddard 2014a, forthcoming). In the NSM framework, as we will see shortly, the conventional understanding of semantics and pragmatics is often turned on its head; or at least, there is an interpenetration of what can be seen in other approaches as two separate fields or perspectives.

Before moving to this, it will be helpful to draw attention to what descriptive linguists have had to say about interjections, which can be roughly defined as fixed, grammatically self-contained utterances. Though interjections are of marginal interest to most linguists, chiefly because they do not combine syntactically with other words (or, at best, only a very limited way), there are some standard formal distinctions drawn between different kinds of interjections. The main division is between so-called primary interjections and secondary interjections (Ameka 1992a, b; Ameka and Wilkins 2006). The former are not identical to other words. Often they have non-word-like phonological characteristics, e.g. English *Ugh!*, *Phew!*, *Mmm!*, which can be sometimes be seen as mimetically motivated; for example, the velar or post-velar fricative in *Ugh*! can be linked with gagging or clearing the throat (Goddard 2014a). Some formally primary interjections, however, do have word-like phonology and may even spawn lexical expressions; for example, the primary interjection *Yuck!* has given rise to the adjective *yucky* and to expressions like *Yuck factor*. So-called secondary interjections, such as *Shit!, Fuck!, Jesus!* and *Christ!*, are identical in form to pre-existing words. The primary/secondary distinction is blurred a little by the existence of modified or sanitised quasi-words, such as, for example, *Geez!* (from *Jesus!*), *Cripes!* (from *Christ!*) and *Darn!* (from *Damn!*). As well, descriptive linguists often use the term interjectional phrase to designate multi-word expressions, such as *My God!*, *Holy shit!*, *God damn it!*.

In this system of nomenclature we can say that *Jesus!* and *Christ!* are secondary interjections, as are *God!, Hell!* and *Damn!* The base words are found in a variety of related interjectional phrases, such as *Jesus Christ*, *Jesus wept!, Thank Christ, For Christ's sake*, *My God!*, and *Go to Hell!* Many of these can augmented, with greater or lesser productivity, with other "intensifier" elements, e.g. in expressions such as *Jesus bloody Christ*, *Fucking Hell*, etc. Recent work in the NSM framework (Goddard 2014a, in press) has suggested that the trichotomy between primary interjections, secondary interjections, and interjectional phrases needs to be elaborated somewhat, but this can be set aside for present purposes.

We turn now to a brief outline of the NSM approach, followed by a review of NSM studies on interjections and swearing.

1.2 The NSM Approach

As is well-known, the NSM research program (Wierzbicka 1996a; Goddard 2011; Goddard and Wierzbicka 2014) gets its name from the metalanguage upon which it is based, i.e. the Natural Semantic Metalanguage. This metalanguage, which has been developed over a decades-long program of conceptual analysis and cross-linguistic empirical research, consists of 65 simple cross-translatable words (semantic primes) and their associated combinatorial syntax. The primes are listed in Table 2, in two versions – English and Spanish. Versions of this metalanguage

Table 2 Semantic primes (English and Spanish), grouped into related categories; after Goddard and Wierzbicka (2014), Travis (2002)

I, YOU, SOMEONE, SOMETHING~THING, PEOPLE, BODY	YO, TU, ALGUIEN~PERSONA, ALGO~COSA, GENTE, CUERPO	Substantives
KIND, PART	TIPO, PARTE	Relational substantives
THIS, THE SAME, OTHER~ELSE	ESTO, LO MISMO, OTRO	Determiners
ONE, TWO, MUCH~MANY, LITTLE~FEW, SOME, ALL	UNO, DOS, MUCHO, POCO, ALGUNOS, TODO	Quantifiers
GOOD, BAD	BUENO, MALO	Evaluators
BIG, SMALL	GRANDE, PEQUEÑO	Descriptors
KNOW, THINK, WANT, FEEL, SEE, HEAR	PENSAR, SABER, QUERER, SENTIR, VER, OÍR	Mental predicates
SAY, WORDS, TRUE	DECIR, PALABRAS, VERDAD	Speech
DO, HAPPEN, MOVE, TOUCH	HACER, PASAR, MOVERSE, TOCAR	Actions, events, movement, contact
BE (SOMEWHERE), THERE IS, BE (SOMEONE/SOMETHING),	ESTAR (EN ALGÚN LUGAR), HAY, SER (ALGUIEN/ALGO),	Location, existence, specification
(SOMETHING) IS (SOMEONE'S)	(ALGO) ES (DE ALGUIEN)	Possession
LIVE, DIE	VIVIR, MORIR	Life and death
WHEN~TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME, MOMENT	CUÁNDO~TIEMPO, AHORA, ANTES, DESPUÉS, MUCHO TIEMPO, POCO TIEMPO, POR UN TIEMPO, MOMENTO	Time
WHERE~PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE	DÓNDE~LUGAR, AQUÍ, ARRIBA, DEBAJO, CERCA, LEJOS, LADO, DENTRO	Space
NOT, MAYBE, CAN, BECAUSE, IF	NO, TAL VEZ, PODER, PORQUE, SI	Logical concepts
VERY, MORE	MUY, MÁS	Intensifier, augmentor
LIKE	COMO	Similarity

Notes: – Primes exist as the meanings of lexical units (not at the level of lexemes) – Exponents of primes may be words, bound morphemes, or phrasemes – They can be formally complex – They can have combinatorial variants or "allolexes" (indicated with ~) – Each prime has well-specified syntactic (combinatorial) properties

have been documented in more than 30 languages from a range of linguistic types and geographical locations. Available evidence indicates that semantic primes can be expressed by words or word-like elements in all or most languages.

Semantic primes are defined as the terminal elements of reductive paraphrase analysis, i.e. meanings that cannot be paraphrased in simpler terms. It follows therefore that to the extent that an analyst can formulate his or her hypotheses in NSM, the analysis avoids implicit definitional circularity and Anglocentrism, two pitfalls which seriously mar much work in semantics and pragmatics. The use of semantic primes also enables an extremely fine-grained resolution of meaning.

[Note[2]] NSM researchers have been co-researching semantics and pragmatics, in an integrated program, for more than 25 years (cf. Wierzbicka 1991, 2006; Goddard and Wierzbicka 2004; Goddard 2009, 2012b; Levisen 2012, among other works). In semantics, the principal mode of representation is known as a semantic explication. A semantic explication is, essentially, a paraphrase framed in simpler words: a way of saying the same thing as the expression being explicated. The main goal of the present study is to arrive at semantic explications of the interjections *Jesus!* and *Christ!*, and to show how corpus techniques can contribute to this goal. (The NSM metalanguage can also be used to formulate cultural scripts, in order to capture norms, standards, and assumptions about ways of speaking that are widely shared within a given speech community.)

Semantic explications are not descriptions of an external observer's point of view, much less an attempt to capture a correspondence between an utterance or expression and a real-world situation. Rather, they are literally paraphrases – models of a speaker's expressed meaning – and, as such, they almost always include a first-person perspective, subjective construals, and expressive/evaluational components. It is notable that included among the 65 semantic primes are I, YOU, PEOPLE and WORD – elements without analogues in other systems of semantic representation (cf. Goddard 2013a, b). These aspects of the NSM approach mean that semantic explications often contain components that would be viewed, in other approaches, as "pragmatic" in nature.

The criteria for a good explication are as follows: (i) it must be well-formed, i.e. be phrased entirely in NSM semantic primes and/or molecules, and conform with the rules of NSM syntax [Note[3]]; (ii) it must be coherent, i.e. make sense as a whole, all anaphoric references must have proper antecedents, temporal expressions must be appropriately anchored, etc.; (iii) it must be substitutable in a broad sense, i.e. be compatible with the range of use of the expression being explicated, generate the correct implications and entailments, and satisfy native speaker intuitions about meaning in context. In relation to the third condition, access to a large body of naturally-occurring examples, whether from a standard corpus or from extensive personal observation is invaluable.

Developing NSM explications is an iterative process. Although there is no mechanical procedure, a heuristic procedure that is often followed by NSM analysts relies on a corpus. In an earlier study, I described the procedure as follows:

---

[2] NSM explications can also include semantic molecules, i.e. non-primitive concepts, definable in terms of semantic primes, that function as intermediate-level semantic building blocks. Although semantic molecules are an important part of the NSM approach, and indeed, are crucial to successfully explicating some lexical domains, they play only a minor role in the current study. Interested readers can consult Goddard (2010, 2011: 375–384, 2012a), Goddard and Wierzbicka (2014).

[3] Also allowed in NSM explications are language-specific morphosyntactic devices that are necessary to implement the valency and complementation possibilities of exponents of primes in a given language. For example, in English NSM the prepositions *to* and *about* are allowed in combinations such as 'do something to something' and 'think about something'.

> The corpus work and the semantic analysis basically worked together as follows. I would provisionally identify a set of examples as likely exemplars of a single semantic category, then draft an initial explication which would make intuitive sense when substituted into these contexts of use (substitutability condition). … After arriving at an apparently satisfactory explication, I would then pull up a second batch of putatively similar examples from the corpus and test the schema against them. Some revision was usually necessary, after which a further set of examples was checked, and so on. This process was carried on iteratively until the schema was proving itself adequate, without revision, against newly selected examples. (Goddard 2007: 121)

Needless to say, this procedure relies on having access to a suitably large and suitably representative corpus, such as are readily available for English, Russian, Chinese, Spanish, and number of other languages. The qualification "suitably large" is an important one. Studies have shown that for fine-grained lexicographical and semantic work, even corpora with words numbering in the tens of millions may not adequately represent lexical phenomena that are quite salient and obvious from a native speaker's point of view (Mair 2007). The qualification "suitably representative" is also important, for certain kinds of language use and language phenomena are likely to be under-represented in standard corpora; for example, intimate talk between lovers; abusive and offensive language; in-group slang and jargon. These and other similar problems are well familiar to those working in corpus linguistics, and various measures can be taken to ameliorate them, including careful use of the "web as corpus", and construction of specialised purpose-built corpora. I will return to these issues below, because both Australian English as a subvariety of English and interjections as a lexical category pose certain problems for standard corpus techniques.

1.3 Approaching Secondary Interjections, NSM Style

The NSM research community has produced a number of significant studies of interjections and "swearing", including Ameka (1992b), Bromhead (2009), Hill (1992), Wierzbicka (1991, 1992, 1996b, 2002), Goddard (2014a, 2014b, in press; cf. Gladkova et al. (Forthcoming)). An early seminal study of swearword interjections (unpublished but available online) was Kidman (1993).

Perhaps the most distinctive aspect of this body of work is showing that it is possible to assign stable, specifiable meanings to interjections, both primary and secondary. It has also demonstrated that what may seem at first blush to be more or less equivalent interjections, e.g. *Yuck!* vs. *Ugh!*, often have significant differences in their meanings and usage patterns, as well as commonalities. A further contribution of the NSM literature on interjections is its cross-linguistic dimension. For example, Wierzbicka (1991) demonstrated the language-specificity even of many primary interjections, by way of contrastive studies of English, Polish, and Russian.

The present study builds on these works, while at same time proposing some analytical innovations. As a preliminary to our treatment of *Christ!* and *Jesus!*,

it will be helpful to first consider another pair of closely-related secondary interjections, namely, *Shit!* and *Fuck!* Explications are presented in [A] and [B] below. It can be seen that each consists of four main sections, which can be labelled Trigger, Reaction, Expressive Impulse, and Word Utterance, followed by a Metalexical Awareness section. This structure or template seems to be shared by many secondary interjections. The two interjections also share several individual components; and others can be seen as variants, with the effect that overall *Fuck!* can be seen as "stronger", both in impact and offensiveness, than *Shit!*

[A] *Shit!*

I know: something happened now in one moment	TRIGGER
I feel something bad because of it	REACTION
I want to say something bad now because of this I want to say it in one moment	EXPRESSIVE IMPULSE
because of this, I say this word {shit}	WORD UTTERANCE

I think about this word like this: "some people can feel something bad when they hear this word some people think like this: "it's bad if someone says this word" "	METALEXICAL AWARENESS

[B] *Fuck!*

I know: something happened now in one moment	TRIGGER
I feel something very bad because of it	REACTION
I want to say something very bad now because of this I want to say it in one moment	EXPRESSIVE IMPULSE
because of this, I say this word {fuck}	WORD UTTERANCE

I think about this word like this: "many people can feel something very bad when they hear this word many people think like this: "it is very bad if someone says this word" "	METALEXICAL AWARENESS

Tracking through the components line-by-line, we can make the following observations. First, the explications share the same Trigger component: 'I know: something happened now in one moment'. The claim is that both *Shit!* and *Fuck!* take off from a recognition or "registration" by the speaker of an immediate event, i.e. the recognition that something has just momentarily happened. The Reaction components are very similar to one another: 'I feel something bad because of it' (for *Shit!*) and 'I feel something very bad because of it' (for *Fuck!*), followed by an Expressive Impulse to say 'something bad' (for *Shit!*) or 'something very bad' (for *Fuck!*), and furthermore 'to say it in one moment'. As one would expect, the Word Utterance component is the same in each case (except, of course, for the choice of the word): a performative use of semantic prime SAY 'because of this, I say this word {XXX}'. Needless to say, it is crucial to the semantic representation of a secondary interjection that what is pronounced is seen as a particular WORD.

Finally comes a section of the explication which can be thought of as "attached" to the word itself: a kind of metapragmatic lexical annotation depicting the speaker's Metalexical Awareness of the status of the word. In the case of *shit* and

fuck, this awareness obviously concerns the word's potentially offensive or "taboo" status. Both are introduced as: 'I think about this word like this: . . .', with the specific content in each case being similar, except that the version for *Fuck!* is "stronger" than for *Shit!* In both cases, there is a pair of parallel components, one phrased in terms of how people can feel when they hear this word ('something bad' for *Shit!*, 'something very bad' for *Fuck!)*, the other in terms of how strongly people can disapprove of someone saying the word: 'it is bad if someone says this word' (for *Shit!*) and 'it is very bad if someone says this word' (for *Fuck!*). The two Metalexical Awareness sections differ in another way as well. For *Shit!* the potential negative reactions are attributed to 'some people', while for *Fuck!* they are attributed to 'many people'. Naturally, the greater potential offensiveness (and potential "shock value") correlates with perceived greater intensity of the word. There is further discussion of the Metalexical Awareness section below.

The most novel aspect of these explications, and the one with the most interesting implications for the relationship between semantics and pragmatics, is the Metalexical Awareness section. To begin with specifics, the point should be made that different degrees of potential offensiveness are modelled here purely in qualitative terms, i.e. without recourse to any numerical scales of offensiveness. Many researchers of offensive language have sought to measure offensiveness in the population at large by way of surveys using rating scales. A five-point scale is typical. Though information gained in this way can be useful, in the NSM view it would make no sense to attribute numerical ratings to the cognitive representations of ordinary speakers. [Note[4]] Nor it is necessary to do so, because, as anticipated by Wierzbicka (2002: 1179–80), it is possible to distinguish a number of qualitatively different categories using combinations such as 'bad' vs. 'very bad', and 'some people' vs. 'many people', as shown in explications [A] and [B]. These combinations, and other similar combinations [Note[5]], are available "free of charge", as it were, as they are independently necessary as part of the NSM.

At a higher level of generality, it might be wondered whether other kinds of words, aside from "swear words", deserve to have a Metalexical Awareness component attached to them; and if so, what range of content and structures are involved. It seems to me that euphemisms presumably involve metapragmatic metalexical components, and that their content would be largely complementary to those found with swear words (cf. Allan and Burridge 1991). Other kinds of

[4] McEnery (2006: 35–36, cf. Note 66, p. 236) makes an interesting observation about the assumptions behind numerical scales, wondering whether it would be fair to assume that the scale is linear, i.e. that there is a steady and even increase in offensiveness as one moves up the scale. For example, would a "bad language word" in category 5 be five-times as offensive as one in category 1? The question highlights the artificiality of taking rating scales literally. Such issues do not arise with subjective-qualitative assessments such as those used in [A] and [B].

[5] For example, words known to be potentially offensive to a minority of people could be tagged with the idea that 'some people, not many people' can react negatively to them. At the other end of the scale, extremely taboo words could be tagged with the idea that 'it is very bad if someone says this word'.

metalexical annotation may be appropriate to model speakers' subjective "knowledge" of sociolinguistic, register, or dialectal associations of words, e.g. that the word *sick* can have the meaning 'very good' in some Australian English "youth speak"; or that *boson* is a "scientific" word.

Finally, what is the significance of presenting the Metalexical Awareness annotation in a different box, offset from the main body of the explication? This presentational device is intended to reflect two things. The first is the intuition that the Metalexical Awareness section is indeed "attached", as it were, to the specific word, rather than forming an integral part of the explication as a whole (the same annotation would occur on all uses of the word *fuck*, for example, including in other swearing contexts, such as the intensifying adjective *fucking* and the "literal" use of the verb *to fuck)*. The second is the hypothesis that individual metalexical annotations can have a "life of their own". It is known that the shock value of particular expressions can decay quite quickly, within decades or so (for example, the words *bloody, damn,* and *fucking* were all once far more offensive than they are today), and yet, there is a functional "pay off" of having strongly offensive terms available for use. This may motivate a speech community to introduce new expressions to fill the slot of a word which is in the process of losing its taboo potency. Given this, it makes sense to see the set of metalexical statuses as somewhat autonomous.

With this by way of background, we are now to addressing the specific research question of this study, i.e. the meaning and usage of the secondary interjections *Jesus!* and *Christ!* in Australian English. Only one more issue remains: how to access a suitably broad sample of naturally-occurring examples?

2 Sources Used in This Study

The present study originated as part of a larger project investigating differences in interactional style between Australian English and American English. For the purpose of the present study, it was decided to restrict our focus of attention to Australian English. Naturally-occurring examples were compiled from two sources: the Australian National Corpus (AusNC) and a selection of six contemporary Australian novels. Though this arrangement was a necessary one in order to acquire a sufficient number of examples, as a spin-off it allows us to compare the pros and cons of using a conventional corpus as opposed to literary material (see Sect. 4).

As noted by a number of researchers, publically available corpora of Australian English are still very limited, compared with what is available for British English and American English (Goddard 2009). This situation has begun to improve since the launch of the Australian National Corpus (AusNC) in March 2012. AusNC [https://www.ausnc.org.au] is an online service whose ultimate objective is to represent the full diversity of language use in Australia, including not only Australian English, but other language varieties as well (Haugh et al. 2009;

Table 3 Number of examples of *Jesus!* and *Christ!* located in post-1985 material in AusNC collections (as at December 2013)

Collection	Description of collection	Tokens of *Jesus!*	Tokens of *Christ!*
Australian Corpus of English (ACE)	1 m words, published text in 500 samples from 15 categories of nonfiction and fiction; from 1986	1	4
International Corpus of English-Australia (ICE-AUS)	1 m words, transcribed spoken and written Australian English; 1992–1995. 500 samples, 60 % speech, 40 % writing [Note[a]]	7	1
		Total: 8	Total: 5

[a]The expression *oh christ* appears once in S2A-026A from ICE, but the speaker was citing, i.e. mentioning, the expression, rather than using it as an interjection

Musgrave 2012). At the time of writing, however, AusNC was still in an early stage of development and it included only nine sub-corpora (termed ‘collections’) of Australian English. The collections comprise several types of data, including literary texts, written correspondence and other documents, transcriptions of audio recordings, and in some collections, original audio and video recordings. Among the collections are two standard corpora of Australian English, Australian Corpus of English (ACE) and International Corpus of English-Australia (ICE-Aus), each about 1 m words in size.

For the purpose of the present study, it was decided to exclude materials earlier than 1985, which meant excluding the COOEE collection of early English in Australia and the AusLit collection of literary works prior to the 1930s. Searches were performed over the remaining collections [Note[6]].

Only tokens of *Jesus!* and *Christ!* as one-word interjections were sought, allowing the occasional combination with another “co-interjection” (e.g. *Oh Jesus*, *Well, Christ!*), but excluding interjectional phrases such as *Jesus Christ, Jesus wept, Christ Almighty, Jesus fucking Christ, Jesus F. Christ, Thank Christ, For Christ's sake* and other variants. Somewhat surprisingly, relevant examples were found in only two collections, as shown in Table 3. Note that ACE and ICE-A include written, as well as (transcribed) spoken, material. The total number of examples was 13 (8 for *Jesus!*, 5 for *Christ!*). Clearly, the AusNC in its present state of development is not an adequate source for this purpose.

As a second source of examples, we located all examples of *Jesus!* and *Christ!* in six contemporary Australian novels, two each by the authors Tim Winton, Christos Tsiolkas, and John Marsden. The publication details are tabulated, along with

[6] Using the AusNC search function [3 Dec 2012], we initially located 189 and 268 files that contained the words *jesus* or *christ*, respectively. We then manually inspected the files to exclude referential uses in religious texts and contexts, and also to exclude false positives turned up by the failure of the search function to identify word boundaries (thus, words like *Christian* and *Christmas* were included in the initial hits for *christ*). The vast majority of the initial hits turned out to be referential uses or false positives.

Table 4 Number of examples of *Jesus!* and *Christ!* located in six Australian novels

Source	Tokens of *Jesus!*	Tokens of *Christ!*
Eyrie. Tim Winton. 2013. Hamish Hamilton. [424 pp]	30	33
The Turning. Tim Winton. 2004. Picador. [317 pp]	18	8
The Slap. Christos Tsiolkas. 2008. Tuskar Rock Press. [483 pp]	4	7
Barracuda. Christos Tsiolkas. 2013. Allen & Unwin. [513 pp]	6	2
Tomorrow, When the War Began. John Marsden. 1993. Houghton Mifflin Harcourt. [288 pp]	2	0
Darkness, Be My Friend. John Marsden. 1996. Pan Macmillan [274 pp]	0	1
	Total: 60	Total: 51

numbers of tokens found, in Table 4. This exercise yielded about 50–60 tokens each of the two interjections.

The selection of the books was somewhat opportunistic, but was guided by the fact the three authors are both well regarded in the literary world and popular with readers. The novels are all set in the present, are realistic in tone, and have a lot of dialogue. Three of them – *The Turning*, *The Slap,* and *Tomorrow, When the War Began* – have been adapted as films or TV series. *The Slap* and *The Turning* are constructed using multiple character viewpoints (actually, *The Turning* is a series of 17 interconnected short stories). Most of the characters in the books are Anglo-Australian, with the exception of *The Slap*, which is explicitly multicultural. The two John Marsden books are from the six-part *Tomorrow* series, which centres on a group of teenagers.

As shown in Table 4, there is great variation between the novels in the number of tokens. My impression is that this is driven primarily by the nature of the characters, but also to some extent by the nature of the story. Winton's 2013 novel, *Eyrie*, which is the top scorer in terms of examples, not only focuses on two Anglo-Australian characters but has more "misery" and interpersonal strife (or so it seems to me) than the other works. It is also notable that in all the literary sources a significant number of tokens are represented as occurring not orally but mentally, i.e. in the thoughts of the characters. To my mind, this unforeseen result raises some interesting general issues for linguistics, but this can be deferred till Sect. 4. Another advantage of the literary material was that it provided rich contextualisation, though some people may query whether this apparent advantage is not offset or even negated by the constructed nature of the material. I will defer this issue also to Sect. 4.

Aggregating the material from all sources gives around 55–65 examples of each interjection, which was deemed to be a sufficient number for the purpose of developing plausible explications. I will designate this purpose-built mini-corpus as the 'Jesus-Christ AustEng Example Set 2014' and in this article refer to it simply as 'the example set'. It is available in full from the author on request.

3 The Meanings of *Jesus!* and *Christ!*

3.1 Overview

Space does not allow me here to detail the process which has culminated in the explications below. It may be of interest to record, however, that the explications are both about "fifth-generation" versions. Broadly, the process followed the steps described in sub-section 1.2; that is, I began with first-draft explications based on intuition, informal observation, and a limited sample of naturally-occurring examples. Over a period of about 3 months, the initial explications were successively revised as they were tested against an increasing number of examples from the example set and other sources, and as I became increasingly aware of ways to improve their coherence and internal logic. In the commentary below I will cite some of the key supporting examples, but I would not want to give the impression that the development process was driven solely by the analytical interaction with the example set. Other source of interpretive ideas included personal observation in daily life, and introspective examination of my own uses of these and other interjections over the period of the main study. I was also approaching the comparison between *Jesus!* and *Christ!* in the context of having recently completed two studies of primary interjections (Goddard 2014a, b, in press), and working concurrently on several other secondary interjections, and on the larger topic of swearing/cursing generally. Developing explications, furthermore, is typically a dialogical process. I was fortunate to be able to discuss and debate aspects of the explications with Anna Wierzbicka, as well as receiving feedback and responses from a symposium presentation (Goddard 2013a, b) of (roughly) third-generation versions.

One initial intuition was that, in Australian speech at least, the interjection *Christ!* sounds "active", and in many uses a bit aggressive or "on edge" (cf. an observation of Wierzbicka 2002, Note 17). *Jesus!*, on the other hand, gives the impression of someone "absorbing a blow". These intuitions suggested that *Christ!* could involve a component expressing the speaker's wanting to do something, while *Jesus!* could involve a component expressing something like helplessness, e.g. 'I can't do anything'. It was also intuitively clear from an early time that *Jesus!* and *Christ!* are more "cognitive" and more complex in their meanings than are *Shit!* and *Fuck!*.

Eventually, I reached the idea that both *Jesus!* and *Christ!* express one's reaction to a thought that has just "hit" one, so to speak. In both explications [C] and [D] below, this Trigger thought is represented as: 'I thought like this now in one moment: "it is like this, I know it"'. As we will see shortly, the implicit information content is sometimes new, i.e. the speaker is having a sudden realization, but one can also utter *Jesus!* or *Christ!* in response to suddenly facing or recalling a known situation.

As well as the shared Trigger component, the two explications share may other significant aspects in their overall structure and content. In fact, they are identical

except for key differences in the Reaction section. For this reason, I will present them both and comment on the shared aspects, before addressing the differences, with reference to examples from the example set. Nevertheless, for ease of reference, the differing components in the two explications are underlined.

[C] *Jesus!* ….

I thought like this now in one moment: "it is like this, I know it now"	TRIGGER
<u>I think about it like this: "this is something very bad for someone"</u> <u>at the same time I think like this: "I can't do anything because of it"</u> I feel something bad because of this	REACTION
I want to say something bad now because of this I want to say it in one moment	EXPRESSIVE IMPULSE
because of this, I say this word {Jesus}	WORD UTTERANCE

I think about this word like this: "some people say this word at many times when they want to say something about God [m] because of this, some people think like this: "it is bad if someone says this word when this someone wants to say something bad" "	METALEXICAL AWARENESS

[D] *Christ!* ….

I thought like this now in one moment: "it is like this, I know it now"	TRIGGER
<u>I think about it like this: "this is something very bad, I don't want this"</u> I feel something bad because of it	REACTION
I want to say something bad now because of this I want to say it in one moment	EXPRESSIVE IMPULSE
because of this, I say this word {Christ}	WORD UTTERANCE

I think about this word like this: "some people say this word at many times when they want to say something about God [m] because of this, some people think like this: "it is bad if someone says this word when this someone wants to say something bad" "	METALEXICAL AWARENESS

Comparing explications [C] and [D], we can see that they share the same Expressive Impulse and Metalexical Awareness sections. But whereas the content of the former is the same as it is with *Shit!* (and presumably with many other "negative" interjections), the content of the Metalexical Awareness section is different. It <u>must</u> be different, of course, because there is nothing inherently offensive, let alone taboo, about saying the words *Jesus* or *Christ*. On the contrary, they could be considered, from the point of view of Christian belief, as very positive – even sacred – words, as names for the Saviour or Messiah. The potentially sensitive aspect of using them as negative interjections (as quasi-swear words) is that such use can be seen as an abuse of their proper, religious, function. Hence, the Metalexical Awareness section has two main ingredients; first, awareness that 'some people' often use this word 'when they want to say something about God [m]' (as indicated by the notation '[m]', the word 'God' is here a semantic molecule); and second, that because of this, 'some people' disapprove of using these special words 'to say something bad'. These components are phrased in such a way that they do not commit the speaker to Christian belief or even to belief in God.

So much for the similarities. In the next two sections we will look into the differences between *Jesus!* vs. *Christ!*, which are located in the Reaction section of the explications.

3.2 Jesus!

According to explication [C], the Reaction expressed by *Jesus!* is:

I think about it like this: "this is something very bad for someone"
at the same time I think about it like this: "I can't do anything because of it"
I feel something bad because of this

The following examples show that *Jesus!* associated with a range of seriously bad feelings such as, roughly speaking, fear, confusion, and distress. Note that in (4) the distressing situation does not directly concern the speaker but an old school friend whose life has fallen apart and who has apparently gone crazy. Examples of this kind account for the fact that the speaker's (purported) thought in the first line of the Reaction section is phrased as: "this is something very bad for someone", rather than "this is something very bad for me". The latter formulation would also not be consistent with the observation that *Jesus!* can be used in response to hearing some terribly worrying news about someone else's health; for example, I myself came out with *Jesus!* several times as a friend told me about his dire situation with bowel cancer.

(1) [. . .] I felt a new kind of fear now; a kind of fear I hadn't even known about before.
'There's nothing', I said to Corrie.
'Oh Jesus,' she said again. Her eyes got very wide and she started going quite white. [Tomorrow, . . ., p?]

(2) [. . .] There was an edge to his voice. 'Man, I don't know what you're talking about.'
Harry said nothing. He watched his employee.
'Jesus, Harry. Are you going to fire me?' The young man's voice cracked and collapsed and he started sobbing. [The Slap, p94]

(3) [. . .] Who's sane and who's crazy? Who's lying and who isn't? Who's seeking attention and who's a bona fide psychotic?
Dean Weily smokes another cigarette, staring from his window at the lights of the city.
"Jesus," he whispers. "This is a bloody nightmare . . ." [ACE 1996]

(4) Good he said. Good. We thought you could accompany him, travel with him up to the city when he goes. You know, a familiar face to smooth the way.
Jesus, I muttered, overcome at the misery and the suddenness of it. I was determined not to cry, or be shrill. [The Turning, p286]

Example (4) illustrates the point, mentioned above, that the speaker's Trigger thought is not necessarily about an event, i.e. about something that has happened. It can be coming face to face with a situation – including a situation that was already known but which the speaker is registering again, in a new context. Example (5) illustrates this too.

(5) But this morning White Point didn't seem like a good idea either. What had he been thinking of? Family? Jesus, there was only Max, his brother, a sister-in-law he'd never met, nieces he'd only heard about [...] [The Turning, p172]

The next two examples show uses that can be seen as "conversational" in nature, in the sense of being highly embedded in the interaction. The speaker is apparently using the utterance of *Jesus!* as a way of displaying a reaction. In (6) the speaker is voicing her disapproval of the tone of the exchange between Carol and Victor. In (7) the speaker's *Jesus!* is displaying the speaker's sudden realization of her own insensitivity at having just mentioned her addressee's deceased father.

(6) Some reunion, said Carol. Actually, it went well, considering.
Kerbside drive-by, I said. The best sort.
No time for arguments. Ideal.
You two, said Vic. Jesus!
Victor, she chided. Not at Christmas.
Let's open another bottle, I said. [The Turning, p212]

(7) So, he said. So. So, how d'you know Faith?
Same way I know your Mum's Doris and your dad's Neville.
He's dead.
Oh. Jesus. Sorry. Fuck. I forgot.
Doesn't matter. [Eyrie, p29]

Finally, we see in (8) an example that could be seen as slightly ironic, or, at least, strategic. The speaker utters *Jesus!* in connection with her reaction to the delicious odours coming from the addressee's kitchen. Her implied message is that the smell is so good that 'it is very bad for someone' (i.e. herself), in that it is impossible for her to resist it.

(8) I could smell it from my place. It's doin me head in.
What is it?
Whatever you're cooking.
You want some?
I've eaten, she said.
Okay.
But, Jesus, she said. Smell's bloody beautiful.
Keely was stumped. She had such an avid look on her face, almost febrile, and she just stood there, as if waiting to be invited in. [Eyrie, p93.]

It should be clear enough from this set of examples of *Jesus!* that the component 'I can't do anything' is a plausible one in all the contexts.

3.3 Christ!

Turning now to *Christ!*, according to explication [D], the Reaction expressed is:

I think about it like this: "this is something very bad, I don't want this"
I feel something bad because of this

Reviewing a selection of examples, one notices that the interjection is often followed by a related utterance with some propositional content. Though the following utterance sometimes describes the bad and unwanted situation as apprehended by the speaker, it more often provides some kind of amplification or explanation of the subject's thought process.

Consider (9)–(11), for example. In (9), the speaker is reacting, disapprovingly, to the sight of wealthy houses on the banks of the Swan River (the interlocutors are in a boat on the river). In (10), the speaker is reacting to the addressee's irritating habit of using the glib expression 'Not a problem'. In (11), *Christ!* appears in a first-person inner narrative.

(9) Hold onto your hat, said Keely, banking across the channel to the next bay. [. . .]
Christ, said Gemma. Look at those houses.
Keely grinned. He thought of Balzac's line – behind every great fortune, a great crime. [Eyrie, p85]

(10) I don't want a bloke anymore, Tom. I haven't got it in me. But I could do with a mate.
Not a problem, he said too lightly.
Christ, will you stop *sayin* that? she said with an exasperated laugh. [Eyrie, p170]

(11) [. . .] but I swung it fractionally to the right and pulled the trigger.
Christ, the noise. It deafened all of us in that confined space. [Darkness, . . ., p239]

It should be clear that the posited Reaction thought 'I think about it like this: "this is something very bad, I don't want this"', is plausible across these three very different situations. It appears that the sub-component 'I don't want this' accounts for the initial intuition that *Christ!* projects something like an "active" attitude on behalf of the speaker. (As mentioned above, I had entertained the hypothesis that *Christ!* included a component such as 'I want to do something (if I can)', but such a component would not be plausible in examples like these, and many others.)

When *Christ!* appears in interactional situations, it often presents a reaction to something that is occurring in the conversational interaction itself. For example, in (11) the speaker is reacting, sarcastically, to her interlocutor Ern's apparent assumption that selling her house would bring in a great deal of money.

(11) 'And settle down amongst all those millionaires,' she said drily. 'Christ, Ern, the sale of my house wouldn't even buy a hut on the peninsula.' [ICE, 1993]

In (12), the topic is the first speaker's unfortunate affair with a sleazy bloke. In saying 'Well, Christ, you were lonely', the second speaker's reaction is not about the affair as such or about the fact that (as she sees it) her friend was lonely at the time. Rather, she is reacting to her perception that her friend is not taking this extenuating circumstance into account, i.e. that she is being too hard on herself.

(12) Maybe. I dunno. I mean it's so grotty. The bloke was the motel manager. He was such a sleaze. I kind of sank into it.
Well, Christ, you were lonely.
Stop defending me. You're worse than Vic. [The Turning, p307]

The next two examples show *Christ!* being employed by the author in depicting the inner thought processes of the protagonist. Using a literary technique that partly merges the perspectives of the character and the narrator, third-person pronouns appear in the depicted thought, but as readers we understand that the thought occurs in the character's mind in a first-person mode, i.e. for (13) 'it's my bloody birthday', for (14) 'What kind of a stiff does that make me?'

(13) She took a quick peek at the TV. The weather report was on and she noticed the date on the bottom of the screen. Christ, she realised, it was *her* bloody birthday today. She could have sworn that she didn't speak out loud but Hugo looked up from the table [. . .] [The Slap, p310]

(14) He knew it was bizarre that he could bear being cuckolded – yes, in time he probably could – but for your wife to think *that* of you? No, he couldn't take it; it was too much. Yet, Christ, what kind of a stiff did that make him? It was sort of funny, in a sick way, and so typical of him. At a time like this, still anxious about his good name. [The Turning, p310]

These examples bring us back to the pros and cons of including literary material in one's sample of examples, an issue that will be taken up shortly. For the moment, however, I would like to conclude the analytical section of this study by claiming to have demonstrated, first, that it is possible to devise explications that plausibly capture invariant meanings for the interjections *Jesus!* and *Christ!* in Australian English; and, second, that although these meanings overlap substantially in several respects, they are far from identical. *Jesus!* and *Christ!* express significantly different (purported) reactions from the speaker.

In broader theoretical perspective, this study challenges a widespread assumption that the meanings of interjections are so context-bound, and their conversational functions are so variable, that they are impervious to semantic analysis. The present study, along with others in the NSM framework, shows that although interjectional meanings involve expressed feelings that are deictically "tied to the moment", these meanings can nonetheless be captured with precision in NSM semantic explications.

It has also been proposed that interjections involving swear words, like *Shit!* and *Fuck!*, or "quasi swear words", like *Jesus!* and *Christ!*, incorporate in their meanings components of ethno metapragmatic knowledge about the status of these particular words in the community of discourse.

4 Discussion: Semantics, Cognition, Corpora

Despite focussing on what could be seen as minute differences between two words, it seems to me that the present study has raised a host of interesting and important issues. In this concluding section I will attempt to identify and express an attitude on some of these issues, in the interests of stimulating debate.

First and foremost, I would like to think that the present study shows how corpora can play an invaluable role in providing evidence for disciplined qualitative

analysis, specifically, NSM semantic analysis (cf. Goddard 2007, 2009; Wierzbicka 2011; Gladkova 2013; among other works). In so saying, I want to react against a view, which I believe is widespread in some quarters, that "real" corpus analysis means quantitative analysis and that there is something disappointing or deficient about research that uses corpora "merely" to locate naturally-occurring examples for qualitative analysis.

Without disputing the value of quantitative analysis for certain purposes, I would like to highlight the value of using corpora as a part of a process of testing and refining semantic hypotheses. Such hypotheses are not "about" corpora as such: they are about how the meanings of words, phrases, and lexicogrammatical constructions can be captured in a rigorous, systematic and formal fashion. Such hypotheses cannot in any sense be "read off" from the data of usage. Rather, they are the product of disciplined conceptual analysis which typically involves introspection, personal observations of language in use, dialogical interactions with fellow language users, language consultants, etc., and a range of logical and theoretical considerations (for example, the internal coherence of an explication, how it relates to other semantically similar and contrasting words, ensuring the simplicity and cross-translatability the metalanguage). Once one is in possession of a viable semantic hypothesis, however, corpora can play an invaluable role in helping to validate, test and refine it. In my view, it would be a healthy thing if linguists working with corpora focussed less on detecting patterns in the data, and more on hypothesis testing.

A very different issue, but one which also has implications for the role of corpora in linguistics, came home forcefully for me as I studied the kind of example material that was available in spoken, as opposed to written, and especially literary, material. It hit me that most linguists, including myself, tend to think of "language" in terms of speech, interaction, communicative writing, and the like. We are not greatly focussed on how words and inner speech are part of our everyday thinking. In the present study, this issue shows itself in the fact that numerous examples of *Christ!* and *Jesus!* (especially the former) are depicted as occurring in the character's minds, and this tallies with my subjective experience, i.e. I can observe myself frequently employing these and other interjections in daily "inner talk".

Perhaps it is time for linguists to open our eyes to the role of words in everyday thinking. There is a certain amount of work in microsociology and phenomenological psychology on this topic. For example, Randall Collins (2004) attempts to explore different modes and forms of thinking (in "Western" life), and, interestingly, he identifies the internal use of cursing, exclamations and other "interaction rituals" as devices for mobilising "emotional energy" (EE) (pp.205–211). [Note[7]] Obviously this is a field of inquiry which poses enormous challenges. At this moment, I would like to make the point, echoing Collins (2004) and the cognitive

[7] Collins (2004), Wiley (1994) and others are very much aware that thinking is often a multi-modal "mash", with visual imagery, sounds, feelings, remembered snatches of conversation, songs, and the like, all competing and/or interacting with vocal self-talk. Inner conversation, furthermore, is often fragmentary and grammatically elliptical. This prompts the thought that interjections, being grammatically self-contained and semantically "of the moment", may be easier to access and compare across individuals than other aspects of vocal thinking.

psychologist Merlin Donald (2001), that certain forms of literature are depictions of the phenomenology of consciousness:

> The best writers have pushed the subjective exploration of the mind much further than would be permissible in clinical or experimental psychology. . . . Novelists in particular often explore our deepest assumptions about awareness. Their portrayals of it constitute a vast, unsystematic collection of phenomena observed from the inside and are possibly the most authoritative descriptions we have. (Donald 2001: 78)

This perspective on literature – i.e. valuing it as a source of insight into subjectivity [Note[8]] – contrasts notably with a prevailing sentiment among many linguists working at the interface of pragmatics and corpus linguistics, which would have it that we should minimise reliance on written material and concentrate on recordings of spoken interactions ("talk", as it is termed in CA jargon). If written material is of any interest, to this way of thinking, the priority would be interactional uses such as email and other electronically mediated communication, rather than on literary forms such as the novel. With Collins, Donald, and others, however, I would like to suggest that literature may provide a valuable point of entry into the "inner" uses of language.

On a more mundane level, I would like to defend the importance of literary material in corpora from two other points of view. This would hardly be necessary if it were not for the reaction one often experiences from linguists to example material that is not sourced from recorded face-to-face interactions. "That's kind of interesting", a typical reaction goes, "but of course it's not authentic (real) data". Against this, I would like to put the position that written uses of language are just as authentic and real, just as much part of the language ecology (in a literacy saturated society such as our own), as are recorded spoken interactions, especially when one considers that reading is just as much a "use" of language as is writing. [Note[9]]

My second point is that although accounts of fictional characters and their interactions are, well, fictional, they contain a great deal of contextualisation which can inform and support interpretation in context. In the novels of Tim Winton and Christos Tsiolkas, for instance, there are many interactions whose fine details of phrasing and expression make sense in view of the reader's insight into the personalities, current mental states, and interpersonal histories of the characters. All this detail is, of course, imagined and constructed by the author, and yet it provides a quality of "meta-data" which would be the envy of empirical researchers in social psychology, conversational analysis, or interactional pragmatics. I would rather retain this material, and handle it with care, than exclude it from consideration as inauthentic.

---

[8] From a slightly different angle, Besemeres (2002), Pavlenko (2006), and Wierzbicka (2014: Ch 18) have argued that some forms of literature, specifically bilingual life narratives, can provide unique access to bilingual consciousness and cognition.

[9] My point is that although occurrences of interjections in literary texts are not examples of interjections being used in real interactions, they are real, authentic examples of how these words are used in the community of discourse and, as such, relevant input for semantic analysis.

Finally I would like to acknowledge that there are many angles (historical, sociolinguistic, cultural) to the use of interjections like *Jesus!* and *Christ!* which are deserving of attention, but which have not been touched upon in this study. It appears to be paradoxically true that although interjections are among the smallest and most marginal parts of any language, they lend themselves to examination from such a multiplicity of perspectives that they seldom disappoint any researcher interested in meaning, history, culture, cognition, and human interaction. We are left, as usual, wanting more.

Acknowledgements The explications were co-developed with Anna Wierzbicka. For helpful comments I would like to thank Bert Peeters and Lara Weinglass. Lara also provided research assistance with the AusNC. Thanks also to Mee Wun Lee for research assistance with the Australian novels. This work was supported in part by the Australian Research Council.

References

Allan, K., & Burridge, K. (1991). *Euphemism and dysphemism. Language used as shield and weapon*. Oxford: Oxford University Press.

Allan, K., & Burridge, K. (2006). *Forbidden words. Taboo and the censoring of language*. Cambridge: Cambridge University Press.

Ameka, F. K. (1992a). Interjections: The universal yet neglected part of speech. *Journal of Pragmatics, 18*(2/3), 101–118.

Ameka, F. K. (ed.). (1992b). Interjections [Special issue]. *Journal of Pragmatics, 18*(2/3).

Ameka, F. K., & Wilkins, D. P. (2006). Interjections. In *Handbook of pragmatics*. Amsterdam: John Benjamins.

Australian Concise Oxford Dictionary. (2004). Melbourne: Oxford University Press.

Beers Fägersten, K. (2012). *Who's Swearing now? The social aspects of conversational swearing*. Newcastle Upon Tyne: Cambridge Scholars.

Besemeres, M. (2002). *Translating one's self: Language and selfhood in cross-cultural autobiography*. Oxford: Peter Lang.

Bromhead, H. (2009). *The reign of truth and faith: Epistemic expressions in 16th and 17th century English*. Berlin: Mouton de Gruyter.

Collins, R. (2004). *Interaction ritual chains*. Princeton: Princeton University Press.

Culpeper, J. (2011). *Impoliteness: Using language to cause offence*. Cambridge: Cambridge University Press.

Donald, M. (2001). *A mind so rare: The evolution of human consciousness*. New York: Farrar Strauss and Giroux.

Gladkova, A. (2013). A cultural semantic and ethnopragmatic analysis of the Russian praise words *molodec* and *umnica* (with reference to English and Chinese). In J. Romero-Trillo (Ed.), *Yearbook of corpus linguistics and pragmatics 2013: New domains and methodologies* (pp. 249–274). Dordrecht: Springer.

Gladkova, A., Vanhatalo, U., & Goddard, C. (Forthcoming). The semantics of interjections: an experimental study with natural semantic metalanguage.

Goddard, C. (Ed.). (2006). *Ethnopragmatics: Understanding discourse in cultural context*. Berlin: Mouton de Gruyter.

Goddard, C. (2007). A "lexicographic portrait" of *forgetting*. In M. Amberber (Ed.), *The language of memory in a cross-linguistic perspective* (pp. 119–137). Amsterdam: John Benjamins.

Goddard, C. (2009). *Not taking yourself too seriously* in Australian English: Semantic explications, cultural scripts, corpus evidence. *Intercultural Pragmatics, 6*(1), 29–53.

Goddard, C. (2011). *Semantic analysis: A practical introduction* (2nd rev. ed.). Oxford: Oxford University Press.

Goddard, C. (2012a). Semantic primes, semantic molecules, semantic templates: Key concepts in the NSM approach to lexical typology. *Linguistics, 50*(3), 711–743.

Goddard, C. (2012b). 'Early interactions' in Australian English, American English, and English English: Cultural differences and cultural scripts. *Journal of Pragmatics, 44*, 1038–1050.

Goddard, C. (2013a). The semantic roots and cultural grounding of "social cognition". *Australian Journal of Linguistics, 33*(3), 245–256.

Goddard, C. (2013b, November). Semantics meets pragmatics (meets sociolinguistics): "Swearing" and "cursing" in Australian English and American English. Keynote address at Pragmatics Meets Semantics Symposium, Griffith University, Brisbane.

Goddard, C. (2014a). Interjections and emotions (with special reference to "surprise" and "disgust"). *Emotion Review 6*(1): 53–63. doi:10.1177/1754073913491843. (Published online 13/09/2013)

Goddard, C. (2014b). On "Disgust". In F. Baider & G. Cislaru (Eds.), *Linguistic approaches to emotions in context* (pp. 73–98). Amsterdam: John Benjamins.

Goddard, C. (Forthcoming). "Swearing" and "cursing" in Australian English and American English. Semantic, pragmatic and cultural perspectives.

Goddard, C. (in press). Ethnopragmatics. In F. Sharifian (Ed.), *Routledge handbook of language and culture*. New York: Routledge.

Goddard, C., & Wierzbicka A. (Eds.). (2004). Cultural scripts [Special Issue]. *Intercultural Pragmatics 1*(2).

Goddard, C., & Wierzbicka, A. (2014). *Words and meanings. Lexical semantics across domains, languages and cultures*. Oxford: Oxford University Press.

Haugh, M., Burridge, K., Mulder, J., & Peters, P. (Eds.). (2009). *Selected proceedings of the 2008 HCSNet workshop on designing the Australian national corpus: Mustering languages*. Somerville: Cascadilla Proceedings Project.

Hill, D. (1992). Imprecatory interjectional expressions: Examples from Australian English. *Journal of Pragmatics, 18*, 209–223.

Holmes, J., & Stubbe, M. (2003). *Power and politeness in the workplace: A sociolinguistic analysis of talk at work*. London: Pearson.

Hughes, G. (1998). *Swearing: A social history of foul language, oaths, and profanities in English*. Oxford: Blackwell.

Jay, T. (1992). *Cursing in America*. Amsterdam: John Benjamins.

Jay, T. (2000). *Why we curse. A Neuro-psycho-social theory of speech*. Amsterdam: John Benjamins.

Jay, T., & Janschewitz, K. (2008). The pragmatics of swearing. *Journal of Politeness Research, 4*, 267–288.

Kidman, A. (1993). *How to do things with four-letter words: A study of the semantics of swearing in Australia*. BA honours thesis. University of New England. http://www.gusworld.com.au/nrc/thesis/intro.htm

Levisen, C. (2012). *Cultural semantics and social cognition. A case study on the Danish universe of meaning*. Berlin: Mouton de Gruyter.

Ljung, M. (2011). *Swearing: A cross-cultural linguistic study*. New York: Palgrave Macmillan.

Longman Dictionary of Contemporary English (LDOCE). (2003). Harlow: Longman.

Mair, C. (2007). Varieties of English around the world: Collocational and cultural profiles. In P. Skandera (Ed.), *Phraseology and culture in English* (pp. 437–468). Berlin: Mouton de Gruyter.

McEnery, T. (2006). *Swearing in English: Bad language, purity and power from 1586 to the present*. London: Routledge.

Mohr, M. (2013). *Holy shit!: A brief history of swearing*. Oxford: Oxford University Press.

Musgrave, S. (2012). An introduction to the Australian national corpus project. *Humanities Australia, 3*, 29–34.

Norrick, N. R. (2009). Interjections as pragmatic markers. *Journal of Pragmatics, 41*, 866–891.

Pavlenko, A. (2006). Bilingual selves. In A. Pavlenko (Ed.), *Bilingual minds: Emotional experience, expression, and representation* (pp. 1–33). Tonawanda: Multilingual Matters.
Stapleton, K. (2010). Swearing. In M. Locher & S. L. Graham (Eds.), *The handbook of pragmatics, Vol. 9 (Interpersonal pragmatics)* (pp. 289–306). Berlin: Mouton de Gruyter.
Travis, C. (2002). La Metalengua Semántica natural: The natural semantic metalanguage of Spanish. In C. Goddard & A. Wierzbicka (Eds.), *Meaning and universal grammar – Theory and empirical findings* (Vol. 1, pp. 173–242). Amsterdam: John Benjamins.
Van Lancker, D., & Cummings, J. L. (1999). Expletives: Neurolinguistic and neurobehavioral perspectives on swearing. *Brain Research. Brain Research Reviews, 31*(1), 83–104.
Wierzbicka, A. (1987). *English speech act verbs: A semantic dictionary*. Sydney: Academic.
Wierzbicka, A. (1991). *Cross-cultural pragmatics: The semantics of human interaction* [expanded 2nd ed., 2003]. Berlin: Mouton de Gruyter.
Wierzbicka, A. (1992). Australian b-words (bloody, bastard, bugger, bullshit): An expression of Australian culture and national character. In André Clas (Ed.), *Le mot, les mots, les bons mots/ Word, words, witty words. A Festschrift for Igor A. Mel'cuk* (pp. 21–38). Montreal: Les Presses de l'Université de Montréal.
Wierzbicka, A. (1996a). *Semantics: Primes and universals*. New York: OUP.
Wierzbicka, A. (1996b). Między modlitwą a przekleństwem: *O Jezu!* i podobne wyrażenia na tle porównawczym [Between praying and swearing: A comparative study of *O Jezu!* 'Jesus!' and other expressions] *Etnolingwistyka* [Ethnolinguistics], *8* 25–39.
Wierzbicka, A. (2002). Australian cultural scripts – *bloody* revisited. *Journal of Pragmatics, 34*, 1167–1209.
Wierzbicka, A. (2006). *English: Meaning and culture*. Oxford: Oxford University Press.
Wierzbicka, A. (2011). *Experience, evidence, sense. The hidden cultural legacy of English*. Oxford: Oxford University Press.
Wierzbicka, A. (2014). *Imprisoned in English. The hazards of English as a default language*. Oxford: Oxford University Press.
Wiley, N. (1994). *The semiotic self*. Cambridge: Polity Press.

A Corpus-Based Analysis of Metaphorical Uses of the High Frequency Noun *Time*: Challenges to Conceptual Metaphor Theory

Shuangling Li

Abstract This chapter attempts to contribute to the ongoing debate over Conceptual Metaphor Theory by investigating its empirical validity in language use with a corpus-based approach. Using the Bank of English (BoE), this study analyses the frequently-occurring linguistic expressions of *time* that are associated with two conceptual metaphors of time (TIME IS MONEY and TIME IS MOTION). The results firstly show that the Lakoffian approach of intuitive metaphor analysis raises questions, as their studies (Lakoff and Johnson, Metaphors we live by. University of Chicago, Chicago, 1980a, J Philos 77(8):453–486, 1980b) fail to mention many frequently-occurring linguistic metaphors of *time* and some of the linguistic examples they gave occur rarely in the BoE. Secondly, the corpus-based analysis reveals more dynamic or complicated linguistic features (e.g. collocational behaviour of certain lexical items and phraseological uses of some linguistic expressions) that cannot be entirely explained or have not been accounted for by the conceptual mapping of Conceptual Metaphor Theory.

Keywords Corpus-based • Metaphorical uses of *time* • Conceptual metaphor theory • Phraseological behaviour

1 Introduction

Conceptual Metaphor Theory (CMT), which proposes that metaphor is not a type of stylistic use of language but primarily a systematic cognitive model of concepts, has given rise to a major revolution in the study of metaphor and consequently CMT is widely applied to metaphor analysis by many cognitive linguists (Lakoff and Johnson 1980a; Lakoff 1993; Kövecses 2010; Gibbs 2011). CMT not only

S. Li (✉)
Department of English, University of Birmingham, Birmingham, UK
e-mail: SXL102@bham.ac.uk

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2,
DOI 10.1007/978-3-319-06007-1_5, © Springer International Publishing Switzerland 2014

highlights the pervasive nature of metaphor in everyday life but also suggests that metaphor influences how people think, speak and act (Lakoff and Johnson 2003; Deignan 2005; Littlemore 2009; Kövecses 2010). More important for linguists is that CMT can be considered to be a conceptual explanation for the metaphorical part of language use.

However, CMT is sometimes criticised in terms of its theoretical assumptions, as in the relationship between metaphor and metonymy (Barcelona 2003); its gradable metaphoricity (Hanks 2006); the explanatory value of conceptual metaphors (McGlone 2007); the classification of metaphor types and constraints on metaphor (Ruiz de Mendoza Ibáñez and Pérez Hernández 2011), or the doubts about its empirical validity (e.g. Deignan 2005; McEnery and Hardie 2012). Nevertheless, the majority of such criticisms have mainly centred on the theoretical aspect and only a few studies have attempted to challenge CMT in terms of the extent to which it can explain real language use, i.e. its empirical validity. Deignan (2005), in one of the most prominent corpus-based studies examining CMT, demonstrates that there are more dynamic and restricted linguistic features to metaphor than CMT suggests.

This study, therefore, aims to further investigate the empirical validity of CMT with a large quantity of attested language (the Bank of English, BoE). This corpus-based analysis focuses on the linguistic metaphors of the word *time* which are associated with TIME IS MONEY and TIME IS MOTION.[1] These two conceptual metaphors are scrutinised because they are the most commonly discussed metaphors of time by Lakoff and Johnson (1980a, 1980b, 2003) where they introduced CMT and illustrated the existence of conceptual metaphor. By analysing in detail the frequently-occurring linguistic metaphors of *time*, this study attempts to explore the metaphorical uses of *time* and more importantly to what extent the corpus data on *time* support or challenge the claims made by CMT.

2 Relevant Terminology in Conceptual Metaphor Theory (CMT)

CMT was first proposed by Lakoff and Johnson (1980a) and further described in Lakoff (1993), Lakoff and Johnson (2003) and Kövecses (2010). It proposes that metaphor in essence is "understanding and experiencing one kind of thing in terms of another" (Lakoff and Johnson 2003: 5). The 'thing' that is to be understood is often an abstract concept which is referred to in CMT as the target domain; the other concrete 'thing' which is used to understand the abstract concept is referred to as the source domain; and this phenomenon of conceptualising one domain in terms of another is called conceptual metaphor (Kövecses 2010). One example frequently

[1] The previous studies also discuss time in terms of other concrete concepts: e.g. TIME IS SPACE, TIME IS AN OBJECT, TIME IS A CONTAINER, TIME IS A CHANGER (Pérez Hernández 2001).

used by CMT is the conceptual metaphor ARGUMENT IS WAR. The usual format of conceptual metaphor is 'A IS B', A being the target domain and B being the source domain. Both the conceptual metaphor and domains are written in capitals.

ARGUMENT IS WAR

Your claims are *indefensible*.
He *attacked every weak point* in my argument.
His criticisms were *right on target*.
I *demolished* his argument.
I've never *won* an argument with him. (Lakoff and Johnson 2003: 4)

As argued by Lakoff and Johnson (1980a, 2003), the existence of this conceptual metaphor is demonstrated by the uses of a large number of relevant linguistic examples (as shown above). Those italicised lexical items such as *indefensible*, *attack* and *won* which are associated with the WAR domain are systematically employed in connection with the ARGUMENT domain (Lakoff and Johnson 2003: 7). These linguistic expressions thus lexically realise the conceptual metaphor ARGUMENT IS WAR and they are referred to as linguistic metaphors. The set of "systematic correspondences" across the two domains (the way elements in the WAR domain correspond to elements in the ARGUMENT domain: e.g. 'the physical attack in a war' corresponds to 'the verbal attack in an argument', and 'win a war' corresponds to 'win an argument') is called "metaphorical mapping" or conceptual mapping (ibid.: 246).

3 Methodology

The corpus being used for this study is the Bank of English[2] (BoE) which is one of the largest general English corpora held at the University of Birmingham, consisting of approximately 450 million tokens. The BoE contains around 70 % of British English, 20 % of American English, and 10 % of other types of English (e.g. Australian and Canadian English). Among the English texts collected for the BoE, about 85 % are written, mainly texts from newspapers and magazines, and the rest 15 % being spoken data. The concordance tool for the BoE is the LookUp software[2] which enables researchers to analyse the corpus data.

The focus of the corpus-based study is on the use of the frequently-occurring linguistic metaphors of *time* which are associated with the two conceptual metaphors (TIME IS MONEY and TIME IS MOTION). The reason for analysing the word *time* as the starting point to approach metaphor lies in the methodological efficiency. This method, compared to other corpus-based methods (e.g. manual

[2] More information about the Bank of English and the LookUp software is available at the website: http://www.titania.bham.ac.uk/

extraction of linguistic metaphors by reading through the corpus), is "superior in terms of data coverage" (Stefanowitsch 2006: 63) because it allows the researcher to focus on the retrieved data of one representative lexical item (*time*) from the target domain (TIME) (see Stefanowitsch and Gries 2006; Rojo and Orts 2010). Additionally, the concept of time is frequently used as the topic to exemplify the existence of conceptual metaphor in CMT, which further indicates the value of analysing the word *time* to evaluate the claims made by CMT.

To assist the extraction of linguistic metaphors of *time*, the study adopted sample analysis as a research strategy. Firstly, a random sample of 500 concordance lines of *time* were selected to identify linguistic expressions which may realise the two conceptual metaphors (based on the definition of metaphor given by CMT: two-domain mappings), and then these linguistic metaphors are further explored in the complete corpus for frequency data and other linguistic features exhibited by these expressions. The main reason for analysing initially with a corpus sample is that even the retrieved data of one item *time* (705,866 instances) are too many to be manually processed for identifying those metaphorical expressions (detecting linguistic metaphors still cannot be fully performed using automated techniques). It may be more practical to conduct the metaphorical analysis of *time* first in a sample. What is more, the results from the initial observation provide a greater focus for further and more detailed corpus analysis.

4 Conceptual Metaphors of Time Realised in the BoE

4.1 TIME IS MONEY & TIME IS MOTION

As demonstrated by Lakoff and Johnson (1980a, 1980b, 2003) (also see Lakoff 1993; Kövecses 2010), time as a more abstract concept is often understood and 'conceptualised' in terms of more concrete concepts, and two common conceptualisations of time discussed in their studies are to perceive time as money and to perceive it as motion.

The conceptual metaphor TIME IS MONEY was first introduced by Lakoff and Johnson (1980a: 7–8). They provided many linguistic examples to illustrate how time is often associated with money (see Table 1). They explained that this everyday association between time and money was based on the cultural background: "work is typically associated with the time it takes" and since "time is precisely quantified, it has become customary to pay people by the hour, week, or year" (Lakoff and Johnson 2003: 8). Also related to this conceptualisation of time – to perceive time in terms of money – is to conceive of time as a limited resource: TIME IS A RESOURCE, or as a valuable commodity: TIME IS A COMMODITY (these two conceptual metaphors will be further discussed in Sect. 4.2). Lakoff and Johnson (2003) state that the concepts of money, resource and commodity are

Table 1 TIME IS MONEY and its associated linguistic metaphors (Lakoff and Johnson 1980a: 7–8, also see Lakoff and Johnson 2003: 7–8)

Conceptual metaphor	Linguistic metaphors
TIME IS MONEY	*You're wasting my time.*
	I don't have the time to give you.
	How do you spend your time these days?
	I've invested a lot of time in her.
	I don't have enough time to spare for that.
	You're running out of time.
	You need to budget your time.
	Put aside some time for ping pong.
	Do you have much time left?
	He's living on borrowed time.
	You don't use your time profitably.
	I lost a lot of time when I got sick.

Table 2 TIME IS MOTION and its associated linguistic metaphors of *time* (Lakoff and Johnson 1980b: 468)

Conceptual metaphor	Linguistic metaphors
TIME IS MOTION	*The time will come when. . .*
	The time has long since gone when. . .
	The time for action has arrived.
	Time flies.

interconnected in our society: "money is a limited resource and limited resources are valuable commodities" (ibid.: 9).

The other conceptual metaphor of time, TIME IS MOTION, reflects the association of time with its movement. Lakoff and Johnson (1980b) gave the following examples of *time* (see Table 2) to show that the perception of time in our daily life and the use of *time* in our language can be grounded in our understandings of motion through space. They suggested that time is often perceived as a moving object where either it is the future moving toward us or it is us who are facing toward the future (ibid.: 468).

4.2 *Frequent Linguistic Metaphors of* Time *in the BoE*

Lakoff and Johnson (1980a, b) used their own linguistic examples (see Tables 1 and 2) to demonstrate the existence of conceptual metaphors and the systematicity of metaphorical concepts (Sect. 4.1). This methodological practice has often been criticised because the conclusions they drew are based on the analysis of invented language examples rather than naturally-occurring data (see Knowles and Moon 2006; Littlemore 2009; McEnery and Hardie 2012). Although Kövecses (2011) defends this intuitive metaphor analysis by explaining that: a. this method saves time; and b. the results generated from this type of analysis coincide with the results from many

Table 3 Frequently-occurring linguistic expressions which are associated with TIME IS MONEY, TIME IS A RESOURCE and TIME IS A COMMODITY in the BoE

	Linguistic expression			Freq.	Freq. per mil.
1. TIME IS MONEY	(1) verbal phrases				
	spend	*time*	v + ing/*with* sb./*on* sth.	12,589	27.98
	waste	*time*	*on*/*in* sth.	3,259	7.24
	make	*time*	*for* sth./*to*-inf.	2,251	5.00
	save	*time*		1,274	2.83
	lose	*time*		1,218	2.71
	(2) nominal phrase				
	a waste of	*time*		1,460	3.24
2. TIME IS A RESOURCE	(1) verbal phrases				
	take	*time*	*to*-inf.	11,879	26.40
	have	*time*	*to*-inf./*for* sth.	10,577	23.50
	give	*time*	*to*-inf.	3,443	7.65
	find	*time*	*to*-inf./*for* sth.	1,746	3.88
	get	*time*	*to*-inf./*for* sth.	1,493	3.32
	(2) other sequences				
	there's/is no	*time*	*to*-inf./*for* sth.	795	1.77
		time	*is/'s/was up*	403	0.90
3. TIME IS A COMMODITY	(1) verbal phrases				
	buy	*time*		745	1.66
	(2) other sequences				
		time	*consuming*	1,647	3.66
		time	*is/'s/was running out*	878	1.95
	running out of	*time*		178	0.40

psycholinguistic experiments, many of the conclusions which have been drawn using this method still seem to be in need of further substantiation.

This study based on the following results further questions the validity of Lakoff and Johnson's method of metaphor analysis, as the list of the frequently-occurring linguistic metaphors found in the corpus seems to be quite different from the list of examples given by Lakoff and Johnson (1980a, b) for these two sets of conceptual metaphors. Furthermore, the corpus investigation reveals more features that have not been considered by their approach.

Table 3 presents the frequent-occurring linguistic metaphors of *time* in the BoE for TIME IS MONEY and its two related conceptual metaphors: TIME IS A RESOURCE and TIME IS A COMMODITY. These linguistic expressions of *time* are loosely classified under one of these three conceptual metaphors, mainly based on the suggestion from Lakoff and Johnson (1980a) concerning the differences between these three related conceptual metaphors and then based on the results of the discussion among a group of researchers. The classification of these linguistic expressions aims primarily to achieve a clear presentation rather than attempt to conceptually categorise them, which is why the fuzzy boundaries between these three conceptual metaphors or mixed mappings exhibited by certain linguistic expressions are not

Table 4 Frequently-occurring linguistic expressions which are associated with TIME IS MOTION in the BoE

	Linguistic expression	Freq.	Freq. per mil.
TIME IS MOTION	*time passed*	985	2.19
	the time has come	949	2.11
	when the time comes	807	1.79
	time went on	563	1.25
	time goes by	354	0.79
	time flies	122	0.27
	time ticked (*away*)	105	0.23
	time slips away/by	45	0.10

further emphasised or focused on in this study. The frequency of the linguistic expressions which contain verbs has also taken into consideration two aspects of the verb-noun collocations: the lemma[3] form of these verbs; and the varied positions of these verbs to co-occur with *time*: allowing zero to two slots between the verb and *time* (e.g. *spend time*, *spend more time*, and *spend too much time*).

Table 4 lists the frequent linguistic expressions of *time* in the BoE which are associated with the other conceptual metaphor: TIME IS MOTION. Similarly, the cases of the linguistic expressions of *time* which involve verb-noun collocations were treated by considering the lemma of these verbs and the flexible positions of the verbs to co-occur with the word *time*.

An initial comparison between these linguistic expressions found in the BoE (Tables 3 and 4) and those examples given by Lakoff and Johnson (1980a, 1980b) (Tables 1 and 2 in Sect. 4.1) seems to show that there is a big difference between the results generated from a corpus-based approach and those from a more traditional approach. Some frequently-occurring expressions found in the corpus are not mentioned by Lakoff and Johnson: e.g. *take time* as the most frequent expression to associate with TIME IS A RESOURCE, and *time passed* as the most frequent expression for TIME IS MOTION. Additionally, some examples they used to illustrate the existence of conceptual metaphors appear to be unnatural. For instance, their invented expression *I don't have the time to give you* (Table 1) cannot be found in the BoE,[4] and the example *budget your time* (Table 1) occurs only once in the 450-million-token corpus.[5]

The comparison between the frequency data for linguistic metaphors in Table 3 and those in Table 4 also indicates that TIME IS MOTION is far less frequently

---

[3] The lemma form of verbs refers to all verbal forms of these lexical items. For example, the lemma form of *spend* (written in capitals as SPEND) includes *spend*, *spends*, *spending* and *spent*. The written form of lemma (e.g. SPEND) is to be distinguished from the capital form of conceptual metaphors (e.g. TIME IS MONEY) and domains (the TIME domain).

[4] The expression *I don't have the time to give you* cannot be found in the BNC (British National Corpus) and in the COCA (Corpus of Contemporary American English).

[5] The expression *budget your time* occurs once in the BNC (100 million tokens) and 6 times in the COCA (450 mil.).

realised than TIME IS MONEY. This result may suggest that time is more commonly conceptualised as money or as a resource than as a moving object. However, this cognitive phenomenon has not been pointed out or proved by previous studies which used the traditional metaphor approach. Furthermore, frequent linguistic metaphors found in the BoE which are associated with TIME IS MOTION do not fully support Lakoff and Johnson's (1980b) proposal of two directions of movement for time (the future is moving toward us; we are facing toward the future; see Sect. 4.1). If time can be conceptualised as a kind of movement as Lakoff and Johnson suggested, it is more likely that time is moving away or more randomly according to the corpus data. Frequently-occurring linguistic expressions which are associated with TIME IS MOTION, such as *time passed*, *time went on* and *time goes by* (see Table 4), may suggest that the concept of time is likely moving away from us; and another three expressions – *time flies*, *time ticked away* and *time slips away* (Table 4) – do not explicitly indicate that time is moving in any particular directions.

Another feature that is shown in the corpus-based analysis which cannot be found using the Lakoffian approach is the phraseological behaviour of the word *time* when it is used metaphorically. As can be seen in Tables 3 and 4, frequently-occurring linguistic metaphors for the two sets of conceptual metaphors are largely verbal phrases or other multi-word sequences. This is also consistent with Deignan's (2005) findings that syntagmatic relation is important to metaphor research. She states that "there are relatively few figurative expressions that appear in isolation, and that the majority form part of a lexical string" (ibid.: 218).

4.3 Linguistic Features That Cannot Be Entirely Explained by CMT

The preliminary presentation of linguistic expressions that are used metaphorically in the BoE (Tables 3 and 4) may seem to support CMT's suggestion that conceptual metaphors can account for a systematic group of linguistic metaphors, although the Lakoffian approach to gathering data raises questions (Sect. 4.2). The following examples in this section will present challenges to CMT as to its claims regarding conceptual mapping and the role of conceptual metaphors in the use of linguistic metaphors.

4.3.1 The Uses of *spend* and *make*

As suggested by CMT, metaphor at the linguistic level is understood as a mapping of lexical items from the source domain to the target domain (Lakoff and Johnson 1980b, 2003; Lakoff 1993). The two verbs *spend* and *make*, which can be perceived as items from the source domain (MONEY), are considered to realise the conceptual mapping of TIME IS MONEY when they are used in connection with the TIME domain. However, the corpus analysis of the uses of *spend* and *make* with

Table 5 Frequent nominal collocates of the lemma SPEND in the BoE (The varied position of the verb has taken into consideration: allowing zero to two slots between the verb and *time*)

	TIME	Freq.	TIME	Freq.	MONEY	Freq.
spend	*time*	12,589	*weeks*	1,354	*money*	5,041
	years	5,852	*minutes*	916	*pounds*	1,879
	hours	3,819	*weekend*	887	*millions*	788
	days	2,701	*Christmas*	547	*dollars*	525
	months	2,273	*morning*	475	*fortune*	495

Table 6 Frequent nominal collocates of the lemma MAKE in the BoE (The varied position of the verb has taken into consideration)

	TIME	Freq.	MONEY	Freq.	Other	Freq.
make	*time*	2,251	*money*	7,785	*difference*	9,569
			profit	2,563	*decision*	8,071
			fortune	1,328	*mistake*	5,639
			pounds	954	*debut*	5,213
			millions	620	*point*	4,627
			dollars	202	*progress*	4,457

time shows that these two verbs exhibit complex linguistic features which are more than simply being the source of evidence for a one-to-one mapping from the source domain to the target domain.

As can be seen in Tables 5 and 6, *spend* and *make* exhibit different tendencies to co-occur with time nouns in the BoE. The verb *spend* occurs more frequently with the nouns which are associated with the TIME domain (e.g. *time*, *years*, *hours* and *days*, see Table 5) than with other nouns which are associated with the MONEY domain (e.g. *money* and *pounds*). The verb *make*, on the other hand, tends to co-occur with the nouns which are associated with the MONEY domain (e.g. *money*, *profit* and *fortune*, see Table 6), and it is even more common for *make* to co-occur with other types of nouns which are not categorised under either the TIME or MONEY domain (e.g. *difference*, *decision* and *mistake*, Table 6).

This different tendency for *spend* and *make* to co-occur with *time*, however, cannot be accounted for by the simple one-to-one conceptual mapping suggested from CMT. Admittedly, it may be possible to partially explain from the cognitive perspective why *make money* is more commonly used than *make time*. Money is normally considered by previous cognitive studies (e.g. Lakoff and Johnson 1980a; Kövecses 2010) as more 'concrete' (physically existing and touchable) than time, and therefore with the verb *make* which is also regarded to associate largely with physical actions, *make money* may correspondingly be perceived as a more acceptable expression than *make time*. Yet the cognitive explanation cannot cover the other linguistic phenomena: for example, why *spend time* occurs more than *spend money*, and why it is even more frequent for *make* to co-occur with *difference* and *decision* than with *time* or *money*. These features of *spend* and *make*, from the corpus-linguistic point of view, simply reflect their unique collocational behaviour: the tendency of a lexical item to co-occur with certain collocates or certain groups of collocates (also see Sinclair 1991, 2004). McCarthy (1990) and Lewis (2000) on the other hand assert that this kind of linguistic phenomena indicates that some collocations can show a certain level of arbitrariness due to the institutionalised feature of language.

Table 7 The pattern 'HAVE *a* ADJ *time*' in the BoE (HAVE is the lemma form of the verb *have*)

HAVE + *a* + adj. + *time*					Freq.
HAVE	*a*	*good*	*time*	/	1,524
		hard		v + ing	988
		great		/	758
		tough		v + ing	306
		difficult		v + ing	273
		wonderful		/	253

4.3.2 The Uses of *have* with *time*

Similarly, based on the suggestion from CMT that lexical items are semantically mapped from the source domain to the target domain (also see Deignan 1997, 2005), it may be assumed that when the verb *have* is used within the TIME domain, it realises the conceptual mapping of TIME IS A RESOURCE. However, some linguistic examples of the co-occurrence of *have* and *time* show that the uses of *have* with *time* can be more phraseological than realising a metaphorical mapping.

For instance, expressions like *have a good time*, *have a hard time* and *have a great time* are different from the verbal phrase *have time*. Each expression is used more as one "unit of meaning" (Sinclair 2004) or as a fixed phrase of discourse functions (e.g. "discoursal expression" (Carter 1998: 67) and "communicative phrase" (Piirainen 2008: 214)) than as a metaphorical expression which reflects the conceptual mapping of MONEY to TIME (cf. a metonymic explanation for *had a great time* in Pérez Hernández (2001)). These expressions fit the pattern 'HAVE *a* ADJ *time*', and the words that appear in the adjectival position mainly denote an evaluative sense (*good*, *hard*, *great* and *tough*, as shown in Table 7). The sequences which fit this pattern may refer to an experience being regarded as good, bad or difficult in the text (see concordance lines 1–5). Those sequences with the adjectives which are associated with negative senses (*hard*, *tough* and *difficult*) are also found in the corpus to show the tendency to precede the gerund, as in *have a hard time understanding* . . . (line 2) and *have a tough time figuring out* . . . (line 4).

1 just basically want them to have a good **time**, and we're teaching our
2 together," says Hill. People have a hard **time** understanding how important a
3 a small party, and everyone had a great **time**. In fact, since Wednesday,
4 can handle it, but they will have a tough **time** figuring out the process in
5 agreement, one would have a difficult **time** coming up with a good

In another expression, *have the time of* one's *life* (see Table 8), the verb *have* cannot be treated as a lexical item being conceptually mapped or metaphorically extended from the RESOURCE domain to the TIME domain. As can be seen in two random instances, lines 6 and 7, the sequence *have the time of their lives* is used

Table 8 The pattern 'HAVE *the time of* poss. *life/lives*' in the BoE ('poss.' is short for possessive pronoun)

HAVE + *the* + *time* + *of* + possessive pronoun + *life/lives*						Freq.
HAVE	*the*	*time*	*of*	*their* *his* *my*	*life/lives*	295

more as a fixed phrase to denote a complete meaning: having or enjoying a great experience.

6 It was one of her last shows and she took them into the studio. They **had the time of their lives**.

7 Besides, John and Carolyn were now **having the time of their lives**. Their faces were on the covers of every top U.S. magazine.

4.3.3 The Uses of *take* with *time*

The uses of the verb *take* with *time* can also exhibit a more phraseological feature rather than being used simply as evidence for the realisation of a conceptual mapping across domains.

The semi-fixed phrase *it takes time*, for example, is used more phraseologically than to just reflect the conceptual metaphor TIME IS A RESOURCE (Sect. 4.2). As a semantic unit, it often denotes a sense that a specific type of work which is to be completed is time-consuming. This phrase is also found in the BoE to frequently co-occur with other items which further contribute to this sense (as shown in Table 9): e.g. adjectives that are related to length: *long* and *some*; modal verbs that imply the long process: *will* and *would*. Additionally, the verbs appear in the to-inf. slot in the sequence *it takes time* to-inf. (*get used to*, *learn*, *build up* and *adjust*) seem to associate more with actions that require a longer time than actions that can be achieved instantly (e.g. *run* or *jump*). Examples of this sequence can be seen in lines 8 and 9.

8 Henry was used as a winger for some of the time with Juve and **it takes time to get used to** reverting back.

9 But such work is necessarily delicate and **it takes time to build up** trust, according to Rennie Johnston, senior outreach worker of long experience.

Another phrase *take your time* (which occurs 353 times in the BoE) is shown to be even more formulaic. Consequently, it seems almost impossible to interpret from this expression the existence of the metaphorical mapping (RESOURCE to TIME). As shown in lines 10–12, *take your time* can be considered as an individual lexical unit which is usually used in a polite way to suggest to the listener that (s)he can do things slowly without hurrying.

Table 9 The uses of the semi-fixed phrase *it takes time* . . . shown in the BoE

The phrase *it takes time*. . .					Freq.
					2,770
(1) co-occurs with adjectives					
it	*took/takes*	*a long*	*time*		625
		some			206
(2) co-occurs with modal verbs					
it	*will*	*take*	*time*		602
	would				
	may				
(3) verbs associated with this phrase					
	it	*takes*	*time*	*to get used to*	268
				to learn	
				to build up	
				to adjust	

10 nobody will be quizzing you, **take your time**." <p> Gradually thing

11 reply, 'answer shortly." You **take your time**, Geo. There's no hurry

12 don't have to call anybody yet. Just **take your time**." She propped her head

4.3.4 The Uses of *time is money*

In the previous three sections, examples were given to question the validity of conceptual mapping and demonstrate that the metaphorical part of language is more complex than mapping lexical items from the source domain to the target domain. The example in this section will illustrate that the suggestion from CMT that conceptual metaphors account for their associated linguistic metaphors may also be problematic.

The phrase *time is money*, which is literally the same as the conceptual metaphor TIME IS MONEY, exhibits unique features rather than simply reflecting the metaphorical mapping. As shown in the concordance lines 13–18, the first three instances (lines 13–15) may be regarded as the conventional metaphorical use of this phrase. However, from the other three concordances (lines 16–18), an extra pragmatic sense can also be detected from the contexts. For example, the co-texts of the phrase: *Digging takes time* (line 16), *if we move too slowly* (line 17) and *wasted time in court means*. . . (line 18), may imply an urge to move fast or take immediate action. This pragmatic sense shown by this phrase in those contexts suggests that the conceptual metaphor cannot entirely account for or predict the use of its associated linguistic metaphors.

13 and process <p> To Americans, **time is money**. We live by schedules, deadlines,

14 Los Angeles, California <p> **Time is money** in any business - and that is

15 the rough-and-tumble 'my **time is money"** world of contemporary American
16 Digging takes time and **time is money.** Even a single trench 1 metre wide
17 says general manager En Kud, **Time is money**, and if we move too slowly,
18 pace at which cases proceed. **Time is money**, and wasted time in court means

As discussed, the phrase *time is money* is used less metaphorically in lines 16–18 than in lines 13–15. This observation could reflect an ongoing process of 'metaphorisation' (Halliday 1994; Deignan 1997) or demetaphorisation, i.e. the ratio of metaphorical or non-metaphorical uses of one linguistic expression among its total uses shifts over time such that this expression may become more metaphorical or less metaphorical. It can also be a phenomenon which is consistent with Hanks's (2006: 17) proposition that "metaphor is gradable". Certain linguistic expressions can exhibit various degrees of metaphoricity in different contexts. However, either of the two explanations would seem to challenge the ideal explanatory role of conceptual metaphor in the uses of linguistic metaphors.

5 Implications of the Corpus Analysis to CMT

5.1 The Systematic Feature of Metaphor

One of the advantages of CMT to the study of metaphor, as pointed out by many linguists (Barcelona 2003; Knowles and Moon 2006; Kövecses 2010), is its suggestion that metaphor is systematic. To some extent, the results from the corpus-based approach support this systematic feature of metaphor, because a group of linguistic metaphors which are cognitively-related can be categorised under one conceptual metaphor: e.g. a group of related verbal phrases *spend time*, *waste time* and *make time* can be considered to be associated with the conceptual metaphor TIME IS MONEY (see Sect. 4.2).

Nevertheless, the corpus analysis also shows that the degree of systematicity raises doubts at the linguistic level. The linguistic expressions which are associated with one conceptual metaphor can exhibit unique collocational or phraseological uses: e.g. the co-occurrence of *have* and *time* can be more phraseological than metaphorical (Sect. 4.3.2). The linguistic expression *time is money*, which is expected to be a 'typical' example for the conceptual metaphor TIME IS MONEY, yet shows different degrees of metaphoricity in different contexts (Sect. 4.3.4). It is, therefore, difficult to determine without specific linguistic contexts which language expressions are associated with one conceptual metaphor.

5.2 *The Model of Conceptual Metaphors and Linguistic Metaphors*

Conceptual metaphor theorists usually present metaphor with a hierarchical construction of a conceptual metaphor and its associated linguistic metaphors (see Sect. 2). However, this framework is largely based on invented language examples, which requires further substantiation with empirical evidence (McEnery and Hardie 2012). Therefore, this study demonstrates that the corpus-based approach can complement the traditional cognitive approach in the process of constructing or developing the conceptual model for metaphor.

For instance, the two conceptual metaphors of time – TIME IS MOTION and TIME IS MONEY – may have been considered by cognitive linguists as equally significant in their conceptual model. The corpus-based approach, on the other hand, reveals that the two conceptual metaphors are by no means realised similarly in the BoE (with TIME IS MOTION far less frequently realised, see Sect. 4.2). Apart from possibly predicting the ratios for different conceptualisations of a concept like time, the corpus-based approach generates more easily the frequency data of linguistic metaphors. These data could highlight the linguistic metaphors that are more likely to be 'typical' or frequent for their associated conceptual metaphor, which to some extent avoids the cases where some linguistic metaphors chosen for the model are rare expressions (e.g. *I don't have the time to give you*, see Sect. 4.2).

5.3 *The Phraseological Feature of the Metaphorical Part of Language*

As shown from the analysis of linguistic expressions of *time* (Sect. 4.2), the two conceptual metaphors are mostly realised by verbal phrases of *time* or other multi-word sequences of *time*, i.e. the metaphorical expressions of *time* are largely phraseological.

This feature of metaphor may have many implications. It first suggests that future research on metaphor should focus more on phrases rather than on single words. Traditional cognitive linguists have mainly studied metaphor based on single lexical items which are used metaphorically. Their lack of investigation of more phraseological expressions may be partly because the traditional approach is not well equipped for such analysis. The concordance tool which is often used in corpus linguistics, for instance, can be extremely useful in the exploration of this part of metaphorical language.

The second application of this feature can be in the teaching of metaphor. Although conceptual metaphors are claimed by theorists to be pervasive in everyday life and valuable to be highlighted to the learners, they are difficult to teach because they are cognitive phenomena and frequently culturally-bound. The finding that

their associated linguistic metaphors are mostly frequently-occurring multi-word sequences or phrases means that teaching these frequent phrases may be a better option. For instance, teaching the verbal phrases like *spend time doing something*, *waste time on something* and *make time for something* would be able to help expand the learners' vocabulary and raise their awareness of the metaphorical conceptualisation of time as money. Additionally, those phraseological expressions, compared to cognitive phenomena (e.g. time being perceived as a moving object), could be more easily included in teaching materials to be used in language classrooms. Again the corpus approach can be of great assistance in the teaching process: providing authentic language contexts for these phrases and revealing their linguistic uses.

Another implication of this finding is for the more quantitative analysis of metaphor. Previous studies of metaphor have been largely restricted to data of a smaller size because identifying metaphor requires a level of manual analysis which is often time-consuming. The phraseological feature of metaphor to some extent provides the potential to identify metaphor automatically or semi-automatically in the corpus, which means that it is possible to investigate the uses of linguistic metaphors with very large quantities of data. A similar suggestion of using phraseological features to identify metaphor is implicit in Deignan's (2005) study which shows that the collocational or patterning features of a lexical item can disambiguate literal and non-literal uses of this item. Hoey's (2005) 'lexical priming' theory also suggests that different uses of a lexical item are primed to show different collocational features. With these theoretical supports, several researchers (e.g. Liang 2013; Patterson 2013) have started to approach automatic or semi-automatic identification of metaphor by looking at different phraseological features of lexical items.

5.4 The Mapping of Lexical Items from the Source Domain to the Target Domain

Implicit in CMT is the suggestion that there is a one-to-one mapping of lexical items from the source domain to the target domain (also see Deignan 1997, 2005). For instance, in the illustration of the famous conceptual metaphor ARGUMENT IS WAR, Lakoff and Johnson (2003: 7) assert that lexical items such as *defend*, *attack* and *win* which are associated with the WAR domain are systematically employed in connection with the ARGUMENT domain. However, based on the results from this study, it can be argued that the one-to-one conceptual mapping is an oversimplified explanation for what is really exhibited by these lexical items at the linguistic level.

As illustrated in Sect. 4.3, the lexical items which are involved in the conceptual mapping can show more dynamic linguistic uses in the target domain. Both *spend* and *make* are considered to realise the mapping of MONEY to TIME when they co-occur with *time*, yet their co-occurrences with *time* cannot be predicted using the conceptual mapping theory (Sect. 4.3.1). Another two verbs *have* and *take* which

are generally held to belong to the RESOURCE domain, when co-occurring with *time*, can form more fixed phrases which exhibit different linguistic uses (Sects. 4.3.2 and 4.3.3).

Deignan's study of animal lexis similarly shows that the metaphorical mapping of lexical items is more complicated than CMT suggests (Deignan 2005: 152–155). The mapping of animal lexis from the source domain to the target domain can take on different grammatical roles. These items which are shown to be predominantly nominal in the source domain seem to take the form of verbs and adjectives when used metaphorically to describe human behaviour and attributes.

Other studies which evaluate conceptual mapping, however, have been conducted from the perspectives which are more theoretical. For example, they may suggest that the mapping should be partial and constrained according to the inherit properties or structures of the domains (see Kövecses 2010; Ruiz de Mendoza Ibáñez and Pérez Hernández 2011). Although these suggestions are undeniably valuable for the development of the framework of CMT, the gap between the theoretical development of the cognitive mapping model and the actual use of language still has not been fully bridged as shown by Deignan (2005) and the results in this study. More work has to be done to further reveal the linguistic features of metaphor which should help promote the development of the cognitive model.

6 Conclusion

This study has provided naturally-occurring linguistic examples as the evidence to challenge the validity of the explanatory role of Conceptual Metaphor Theory (CMT) in language use. Firstly, the intuitive metaphor analysis used by traditional conceptual metaphor theorists raises questions. The list of the frequently-occurring linguistic metaphors of *time* found in the BoE appears to be different from the list of language examples given by Lakoff and Johnson (1980a, b). Secondly, CMT does not seem to be able to account for more complex features of language. For instance, CMT may suggest the conceptual mapping of MONEY onto TIME as the explanation for *spend time* and *make time* to be used metaphorically, but it cannot fully account for the unique collocational behaviour of *spend* and *make*. Similarly, CMT suggests a simple mapping of lexical items from the source domain to the target domain to explain the semantically 'transferred' use of *have* and *take* in the TIME domain. However, the corpus data would seem to show that the co-occurrence of either verb (*have* or *take*) and *time* can be used more phraseologically than exhibiting the metaphorical mapping (e.g. *have the time of their lives* and *take your time*).

Acknowledgments The author is greatly indebted to Dr. Crayton Walker for his guidance and the two anonymous referees for their comments. Needless to say, the author is responsible for any remaining errors.

References

Barcelona, A. (Ed.). (2003). *Metaphor and metonymy at the crossroads: A cognitive perspective*. Berlin: Mouton de Gruyter.

Carter, R. (1998). *Vocabulary: Applied linguistic perspectives* (2nd ed.). London/New York: Routledge.

Deignan, A. (1997). *A corpus-based study of some linguistic features of metaphor*. PhD dissertation. The University of Birmingham, Birmingham.

Deignan, A. (2005). *Metaphor and corpus linguistics*. Amsterdam: John Benjamins.

Gibbs, R. (2011). Evaluating conceptual metaphor theory. *Discourse Processes, 48*(8), 529–562.

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). Oxford: Blackwell.

Hanks, P. (2006). Metaphoricity is gradable. In A. Stefanowitsch & S. T. Gries (Eds.), *Corpus-based approaches to metaphor and metonymy* (pp. 17–35). Berlin: Mouton de Gruyter.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London/New York: Routledge.

Knowles, M., & Moon, R. (2006). *Introducing metaphor*. London/New York: Routledge.

Kövecses, Z. (2010). *Metaphor: A practical introduction*. Oxford: Oxford University Press.

Kövecses, Z. (2011). Methodological issues in conceptual metaphor theory. In S. Handl & H. Schmid (Eds.), *Windows to the mind: Metaphor, metonymy and conceptual blending* (pp. 23–40). Berlin: Mouton de Gruyter.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 202–251). Cambridge: Cambridge University Press.

Lakoff, G., & Johnson, M. (1980a). *Metaphors we live by*. Chicago: University of Chicago.

Lakoff, G., & Johnson, M. (1980b). Conceptual metaphor in everyday language. *The Journal of Philosophy, 77*(8), 453–486.

Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. London: The University of Chicago Press.

Lewis, M. (Ed.). (2000). *Teaching collocation: Further developments in the lexical approach*. Boston: Thomson Heinle.

Liang, M. (2013, July). *Sparing a free hand: context-based automatic categorisation of concordance lines*. Paper presented at the 7th International Corpus Linguistics Conference (CL2013), The Lancaster University, Lancaster.

Littlemore, J. (2009). *Applying cognitive linguistics to second language learning and teaching*. Basingstoke: Palgrave MacMillan.

McCarthy, M. (1990). *Vocabulary*. Oxford: Oxford University Press.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

McGlone, M. (2007). What is the explanatory value of a conceptual metaphor? *Language & Communication, 27*, 109–126.

Patterson, K. (2013, July). *The identification of metaphor using corpus methods: Can a re-classification of metaphoric language help our understanding of metaphor usage and comprehension?* Paper presented at the 7th International Corpus Linguistics Conference (CL2013), The Lancaster University, Lancaster.

Pérez Hernández, L. (2001). Metaphor-based cluster models and conceptual interaction: The case of 'time'. *Atlantis, 23*(2), 65–82.

Piirainen, E. (2008). Figurative phraseology and culture. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 207–228). Philadelphia: John Benjamins Publishing Company.

Rojo, L. A., & Orts, M. A. (2010). Metaphorical pattern analysis in financial texts: Framing the crisis in positive or negative metaphorical terms. *Journal of Pragmatics, 42*, 3300–3313.

Ruiz de Mendoza Ibáñez, F. J., & Pérez Hernández, L. (2011). The contemporary theory of metaphor: Myths, developments and challenges. *Metaphor and Symbol, 26*(3), 161–185.

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. M. (2004). *Trust the text: Language, corpus and discourse*. London/New York: Routledge.

Stefanowitsch, A. (2006). Words and their metaphors: A corpus-based approach. In A. Stefanowitsch & S. T. Gries (Eds.), *Corpus-based approaches to metaphor and metonymy* (pp. 63–105). Berlin: Mouton de Gruyter.

Stefanowitsch, A., & Gries, S. T. (Eds.). (2006). *Corpus-based approaches to metaphor and metonymy*. Berlin: Mouton de Gruyter.

Part II
Current Approaches to the Pragmatics of Culture and Society

Horace, Colors, and Pragmatics

Jacob L. Mey

Abstract When discussing the use of color terms in societies that diverge from ours (in time, space, and other respects), the problem is not only that the ways colors are perceived, differ, but also that the use and importance of colors in the daily lives of users vary vastly from society to society. (The (in)famous, now much critiqued case is of course that of the many words for 'snow' in the Inuit languages and dialects).

The chapter sets out to examine the null-hypothesis ('Horace was suffering from color blindness') and finds there is no evidence to support it in the corpus examined here (comprising the bulk of his poetic production, with the exception of parts where color terms are less likely to occur: most of the *Sermones*, the *Epistulae* and *Ars Poetica*). The pragmatic angle on all of this is that one cannot discuss the use of language in the abstract (e.g. based on isolated vocabulary entries). What is needed is to place the study in a wider, societal context, to the extent that this is possible (avowedly, there are difficulties in cases like Horace's, where the relevant societal structures have been changed or lost over time, such as it has happened to the 'languaculture' of which Horace was a member). Even so, the poet emerges from my study as one who decidedly has a certain 'feel' for color, but perhaps did not always use it in ways that we consider familiar.

Keywords Q. Horatius Flaccus • Color terms • Color relativity • Historical pragmatics • Latin lexicon & phraseology • Mimesis • Metaphor & metonymy

J.L. Mey (✉)
Institute of Language and Communication, University of Southern Denmark, Odense, Denmark
e-mail: jam@sdu.dk

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2,
DOI 10.1007/978-3-319-06007-1_6, © Springer International Publishing Switzerland 2014

1 Introduction

When discussing the use of color terms in societies that diverge from ours (in time, space, and other respects), the problem is not only that the ways colors are perceived, differ, but also that the use and importance of colors in the daily lives of users vary vastly from society to society. (The (in)famous, now much critiqued case is of course that of the many words for ‘snow’ in the Inuit languages and dialects).

Looking at colors through the prism of history, one finds remarkable differences of use. Considering how a Roman poet of the Augustean age like Horace uses colors, one might even be tempted to raise the question if the poet was basically color-blind: his palette is in many ways much poorer than the one we modern, Western users, employ in our daily practice. Thus, the color ‘blue’ (Latin *caeruleus*, lit. ‘the color of the sky’) is almost totally absent, while the use of ‘standardized’, stereotyped colors, such as ‘green’ (in its various lexical realizations) abounds.

The chapter sets out to examine the null-hypothesis (‘Horace was suffering from color blindness’) and finds there is no evidence to support it in the corpus examined here (comprising the bulk of his poetic production, with the exception of parts where color terms are less likely to occur: most of the *Sermones*, the *Epistulae* and *Ars Poetica*). Thus I land on the alternative hypothesis, namely that Horace uses colors in different ways than we do; this hypothesis seems to be borne out by the examination of the various instances of color uses, arranged according to different classifications.

The pragmatic angle on all of this is that one cannot discuss the use of language in the abstract (e.g. based on isolated vocabulary entries). What is needed is to place the study in a wider, societal context, to the extent that this is possible (avowedly, there are difficulties in cases like Horace’s, where the relevant societal structures have been changed or lost over time, such as it has happened to the ‘languaculture’ of which Horace was a member). Even so, the poet emerges from my study as one who decidedly has a certain ‘feel’ for color, but perhaps did not always use it in ways that we consider familiar.

More generally, our interest in what has been called ‘historical pragmatics’ should not limit itself to merely recording some obsolete uses of language (e.g. in dealing with speech acts and phraseology); it has to understand the linguistic practices of earlier times on their own terms, that is, by placing the texts in their historical *contexts*. The much maligned science of what the German philologists used to call *realia*, in the tradition established by Ulrich von Wilamowitz-Möllendorf and others in works entitled ‘Reallexika’ may yet experience a come-back, provided it is practiced in the sense defined by Jucker & Taavitsainen as: “a diachronic study of language use and human interaction” that “also includes the *social context* of language use” (2010: 5; my emphasis).

2 The Eye of the Beholder...

It is a well-known fact (attested by experience and by the words of the literati) that 'beauty is in the eye of the beholder' (subtext: and not just in the object beheld—on the beauty of which there may be indeed be as many opinions as there are heads and hearts, as in the case of the love attributed to the latter by Anna Karenina; Tolstoy 1952: 155–156).[1]

Something similar seems to be happening when we start looking at the ways various people(s) perceive colors. When I first lived in Tokyo, I was struck by the particular color of the traffic lights we normally associate with the color 'green': the Japanese lights were more like what we would call 'blue'. This creates normally no problem for the Japanese while in Japan, but if they have to name the color of the traffic light in English, there is confusion. By contrast, the word for 'green' in Japanese, *midori,* does not at all fit the color of the signal for 'GO'; rather, it denotes a color we associate with 'green', as in the (default) color of grass. For the color of the traffic light 'green', Japanese has a particular word, *ao* (sometimes described, in purported accordance with the shape of its *kanji*, as "the color of [young] plants in moonlight" (O'Neill 1982: 68, *s.v.*).

A practical, and indeed *pragmatic* aspect of this naming is the fact that Japanese users taking a driver's test outside of Japan often fail, because they name the 'GO' signal 'blue' rather than 'green'. Over the years, I have noticed how the Japanese authorities gradually replaced the *ao* lights with ones that are more like *midori*, in an effort to end the confusion; but the Japanese continue to call their 'GO' signal *ao*, no matter what the bureaucrats at the Ministry of Transportation decree![2]

At the other end of the globe, in Ireland, a similar 'confusion' seems to take place. From the times I shared a house with my Irish-speaking old friend Mártín Ó Murchadha (Martin Murphy), who later became a professor at University College Cork, I remember the endless discussions we had about the color of the sea, the skies, grass, and so on. Martin was adamant in calling the colors he saw in these objects *glas*, irrespective of their 'real' color. I asked him: "But Martin, if the sea is blue, what do you call it?" "*Glas*", he answered. "But what if it's gray?" "*Glas*". "But if it is neither gray nor blue, but green?" "It still is *glas*", Martin said. At which point I gave up trying to understand the mysteries of this color word, but it comforts me a little when I notice that even an accomplished linguist like the late Louis Hjelmslev, in his *Prolegomena*, when taking pains to dissect this Gordian knot, does not go further than merely describing the area of *glas* in Celtic as corresponding to the three colors 'green', 'blue', and 'gray' in other languages like Danish (1943 [1993]: 48–49), adding laconically that the color spectrum is

---

[1] An early version of the saying is due to Shakespeare, *Love's labour's lost*:
"Beauty is bought by judgement of the eye." (Act 2, Scene 1).

[2] The traffic lights in the Kansai area (cities like Osaka, Kyoto, Kobe and so on) always have seemed to me to be more in accordance with Western practice—I have no explanation for this phenomenon, if indeed it exists.

'amorphous' in principle, and stays that way until the individual languages operate their different 'cuts' in accordance with the ways they "form" its purport—a view that is as far removed from a pragmatic conception as possible.

3 A Pragmatic View of Colors

For a pragmatic linguist, when faced with a problem like the above, the first question to ask is always: How does this matter relate to the user of language? For many linguists, including the earlier quoted Hjelmslev, this was not a legitimate question within linguistics proper (the "immanent" kind, as Hjelmslev was wont to call it). Saussure had taught us that the way we 'associate' meanings with sounds is arbitrary, and that in particular our usage of words is so personalized and psychologically opaque that no true scientist would want to get his or her hands dirty dealing with it. It took some anthropologists (especially the Californians Brent Berlin and Paul Kay in the late sixties of the past century; Berlin and Kay 1991 [1969]) to put the problem of 'color relativity' back on the linguists' agenda; but also their work was mainly selective and descriptive, looking at instances of particular uses.

In my own handling of the color continuum, I intend to make the pragmatic point of view a bit clearer. It is indeed the case that we use colors idiosyncratically, both personally and nationally; but even so, there are regularities, and they may have something to do with the *things* we do (even that old stalwart, the Eskimo word for 'snow', may be brought into the arena at this point, albeit without overly relying on the often anecdotal and unreliable figures that are quoted in his connection; cf. Pullum 1992: 159–171). In the sequel, by following a particular user, the Roman poet Q. Horatius Flaccus, on his way through the color world as it is reflected in his writings, and paying attention to his use of the Latin color terms, we will notice how indeed a profile emerges from this exercise.

But let me first say a few words about how I came to be interested in this remote and rather murky nook of Classical philology in the first place.

4 Me and Horace

I got the idea for this article by reading Horace. Not just once, but numerous times over the years, ever since I discovered him, back in the forties of the last century, attending a Jesuit High School, where he was regarded with some suspicion, especially by our Latin teacher (a Jesuit himself), whose competence in Latin was rivaled by his moral scruples. As a result, we were only allowed to read carefully 'castigated' editions, thus being deprived from enjoying some of the finest verse in Latin poetry.

Later on, when I taught Latin (substituting Horace for the all-too-easy and pedestrian C. Nepos) to some junior Jesuits who already had a solid foundation in the language, I came up against the same bowdlerizing, restrictive editing and reading policies. Somehow, I managed to obtain a more or less complete edition of the poet's work, from which I carefully culled some not-too offensive passages, so I could show those young monks at least the beginning of the road to Latin perfection and moral perdition.

Ultimately, and much later, it fell into my lot to teach an advanced Latin class in my own university, called 'Latin Conversation'. Since the students were more or less tongue-tied in Cicero's beautiful idiom, I decided to loosen them up a little bit by bringing to class some nice extracts from Horace for discussion and conversation. I made up what I hoped would be attractive thematic headings, such as 'Horace and Women' (a very rich repository of juicy tidbits), 'Horace and Men' (a little less extensive, but no less interesting), 'Horace and Wine' (this implied going to the 'sources' and figuring out what kind of wines still might be available and/or drinkable, such as the Falerno, to-day classified, with the epithet 'del Massico', as a 'DOC' and currently on its way to a three-star grade; cf. Johnson 2009)[3]; and finally 'Horace and Colors', which came to be the direct inspiration for my current piece.

The way I went about my research in connection with this teaching was definitely impressionistic; but even as impressions go, I ended up with some rather strong notions. For one thing, Horace did not seem to have consciously 'colored' his verse; in any case, his palette looked rather limited to me. So the thought struck me that he might have been color-blind—which was actually my original assumption, when I started to revamp my original notes for this article. There, the principal (and truly transcendental) problem turned out to be, of course, what it would take to determine color-blindness in a person over 2,000 years dead.

5 A Color-Blind Poet? A First Approach

What at first blush seemed to be a straightforward query turned out to be, at closer quarters, a potential quagmire of unwarranted assumptions and interculturally (not to omit cross-era-wise) skewed fits. Even assuming that color-blindness did exist in Horace's times (and there seems to be no reason to doubt that), the question still remains if the people of those times considered such a 'blindness' as a deficit in humans, or even whether they were at all aware of the phenomenon.

Discussions of this kind are intimately related to the discourse on how we (mentally and physically) represent reality. For the classical authors, as for their contemporaries, this question always was a matter of what Plato used to call 'imitation' (*mímesis*), a question of being true (or truer) to life. Clearly, in painting,

[3] ". . . the best-known wine of ancient times, probably white and sweet. Today elegant and red." (Johnson 2009: 107).

colors played a major role, inasmuch as they helped create that sensation of faux reality for which the best classical painters were famous (the story about the contest between Zeuxis and Apelles may serve as an example).[4] However, when it came to describing reality in words, the contribution of the colors to the goals of such 'realism' did not really seem all that important. On the contrary, the use of colors and color imagery seemed to be related, to a large extent, to metaphoric and metonymic usage of color terms; alternatively it could even function as a form of 'social commentary' (the expression is Clarke's 2003: 299).[5] The tendency to consider such 'derivative' uses of color terms as primary, or at least as more important when it comes to doing literary studies, is prevalent even in our times, as we will see in the following.

Clarke (2003: 6) refers to earlier work, in particular a study by Edgeworth (1992), who divides the literary usage of color terms into six categories:

- *formulaic* (repeat phrases, such as the famous *flavus Tiberis*, the 'yellow Tiber', *passim* in Horace; see below)
- *functional* (expressing some important aspect of the narrative)
- *allusive* (referring to usage by earlier authors)
- *decorative* (adding an embellishing detail)
- *cumulative* (when color terms appear in clusters; Edgeworth's own work on Virgil (1992) makes extensive usage of this type of occurrence, which, however, seems of minor importance in the study of Horace's works)
- *associative* (linking together episodes in the work).

6 Nature's Picture, Courtesy of Horace

None of the above-listed categories seem apt to describe what I would call a 'simplistic-naturalistic' use of color terms, viz., when they are employed to describe what is 'out there', without any allusions or associations either to occurrences within the text or to outside literary authorities or conventions. What I'll be primarily looking for is the way Horace 'paints', in words of color and colorful pictures, the things he experiences, using his sense of vision and other senses. Secondarily only, I'll touch upon what these colors may 'mean' in a derived sense,

[4] The two Sicilian master painters, Apelles and Zeuxis, are reported to have engaged in a contest for the title of 'Supreme Painter'. The criterion was naturalness. Apelles painted a *nature morte* with fruits, done so well that one member of the jury was tempted to grab a fig and taste it—he only got his hands dirty, of course. But Zeuxis was even more proficient: he painted a floral composition that was so true to nature that a bee mistook the flowers for real and got stuck in the wet paint. (As recounted by the Greek philosopher-linguist Diodorus Siculus).

[5] One should also remember, as Clarke remarks, that "the Roman response to certain colours and colour terms may have been different from our own" (2003: 3), echoing a sentiment earlier expressed by Alice Kober (1934). On this, see also Steinmayer's and others' (2000) discussion on the Classics website.

that is, metaphorically or even metonymically, as this aspect of the use of colors has been covered admirably by authors such as Nisbet and Hubbard (1970, 1978), Garrison (1991), and more recently Clarke (2003).

To illustrate the difference between these two approaches, compare, as contrasting instances, the use of *candidus,* 'white' in the famous ode depicting Mount Soracte in winter:

Vides ut alta stet nive candidum
Soracte ... (*Od.* I, 9, 1–2),
("You see how Soracte stands white in its depth of snow")

With the use of the same adjective, *candidus*, in a quite different setting (corresponding to the other approach), as when Horace speaks of the 'honest mind', *candidum ingenium*, and of a 'dazzling girl', *puellae candidae*, in one of the *Epodes* (11, 11; 11, 27); here, the color terms carry a distinctly metaphorical flavor.

Similarly, among the occurrences of terms for 'black', we find a frequent term, *niger*, literally depicting a black state of affairs; here, forests, hills, smoke, waters, foliage, hair, eyes, teeth, even nails, etc. are called black (for an extensive listing, see below, Sect. 7.3). We may then compare this usage with the use of the same word for the sun, for fires, for the black netherworld (Orcus), where 'Mercury's black herd' (i.e. the ghosts of the dead; *Od.* I, 24, 18) reside; also for Eurus (the 'black' East wind; *Epod.* 10, 5), Meriones (called a 'black' person, perhaps on account of his deeds; *Od.* I, 6, 15), and so on. Compare also the near-exclusively metaphorical use of another of the words for 'black', *ater*, as in *atra cura* 'black worry' (*Od.* III, 1, 40), *morti atrae* 'to black Death' (*Od.* I, 27, 13), *atra fila* 'the black threads [spun by Fate]' (*Od.* II, 3, 16), and so on.[6]

It is my thesis that Horace, in his descriptive (corresponding more or less to Edgeworth's 'decorative') use of color terms employs a rather meager palette, especially in his descriptions of nature (which, as we will see, revolve around the key terms of 'black' and 'white'). While theoretically, this could be attributed to a lack of *perception* on the part of the poet, the jejune expressions could equally well be due to a corresponding lack of *need*.

Under this interpretation, 'painting' in words is thought of as secondary to the painters' creation of real pictures; the primary aim of the poetic 'painting' not being the representation of some natural state of affairs, but to steer the recipients' perceptions and emotions in a certain, preferred direction. In this primitive *ars poetica* of verbal painting, the color words as such did not yet fulfill the all-important role that they would come to play much later, when the poets of the nineteenth and twentieth centuries sought to emulate their colleagues, the painters, by expressing themselves directly in the language of colors and related expressions.[7]

---

[6] It is by no means a coincidence that the illustrative examples here almost spontaneously came out in 'black and white'; more on this later.

[7] The French poet Arhur Rimbaud's (1854–1891) sonnet *Voyelles* ("A noir, E blanc, I rouge, U vert, O bleu: voyelles, ...") is the classical example of this trend.

Even with these reservations and caveats (and leaving aside the question of a possible physical human condition), we may gather some evidence from Horace's works for what one could call his '*literary*' color-blindness. In his description of e.g. the sea, or a winter's day in the mountains, our poet focuses on immediately available contrasts (primarily between black and white: compare the earlier quoted *nive candidum Soracte* and *ater sinus Hadriae*); as will emerge from the discussion of the data below, this is a pervasive, not an isolated phenomenon. But even in the poet's 'tropical' (mainly metaphorical or metonymical) uses of color, the predominance of these contrasting 'non-colors' is evident: whereas the 'whites' are used to express pleasant feelings or positive traits, as in *nota cressa* 'a white [i.e. positive] mark' on the calendar (*Od.* I, 35, 10), the 'blacks' are prevalently related to more negative surroundings (cf. *atra cura* 'black Worry' (personified as a goddess; *Od.* III, 1, 40)).[8]

In the following sections, I will present the corpus on which I build my hypotheses.

7 The Data

7.1 Introduction

The following is based on a lexical breakdown of Horace's verse, mainly the *Odes* and *Epodes* (in addition, the *Carmen Saeculare* and a single one of the *Sermones* (I, 9) have been scanned for color terms).

I have looked for lexical items having to do with color. These were mostly adjectives, such as the equivalents of English 'white', 'black', and so on; occasionally, verbs (mostly inchoative) and other (mostly derived) items have been taken in as well (English equivalents of these are words like 'redden' or 'incandescent').

The 139 lexical occurrences are listed alphabetically in their respective categories, each with a minimal context, so that their proper semantic value becomes clear; sometimes, the contexts have been expanded for better understanding. All terms have been translated and/or paraphrased.

Following the initial listing, the individual terms have been grouped into these major functional categories:

Nature
Body
Objects
Metaphor
Unspecified.

[8] Compare that even in the Bible, 'black' carries these connotations. The Bride in *Canticle* defiantly proclaims her beauty *despite* her blackness:

"I am black but comely, ye daughters of Jerusalem" (*Cant.* 1:5).

7.2 Color Terms in Horace

There follows a listing of the color terms and color related words in the Horace corpus on which this study is based.

albus, albicare, albescere 'white, be/become white'
ardens 'burning, red-hot'
ater 'black'
aureus 'golden'
caeruleus 'blue'
candidus, candens 'white, gleaming'
canus, canities 'white(ness)' (orig. of hair)
cereus 'wax-colored'
coruscus '(red-)flickering'
cressus 'white' ('color of chalk')
eburnus 'white' ('ivory-colored')
flavus 'reddish-yellow, blond, auburn'
fulvus 'brownish yellow'
furvus 'dark-colored'
igneus 'fiery-red, hot'
lividus '(blue-)black'
luridus 'greenish-yellow'
luteus 'yellow(ish)'
niger 'black'
nitor, niteo, nitidus 'shine/ing'
murreus 'dark golden'
niveus 'snow-white'
pallidus, pallor 'pale(ness)'
pullus 'dark green'
purpureus 'purple'
ravus 'yellowish gray'
roseus 'pink'
ruber, rubens, erubescere 'red, turn red'
stercor 'brown (color of crocodile) manure'
viola '[color of the] violet (purple or yellowish)'
viridis, virens 'green, fresh'
vitreus 'color of glass, resplendent'

7.3 A Breakdown of 'Color collocations' by Categories

7.3.1 Geography & Nature

albus: albus Notus 'the white South wind' (maybe 'clearing the skies'; the Greeks' *nótos*, our scirocco or mistral) (*Od.* I, 7, 15–16)

albus: stella alba 'a white, i.e. dazzling star' (*Od*. I, 12, 27–28)
albus: alba populus 'the white poplar' (*Od*. II, 3, 9)
albus: in album alitem 'into a white bird' (most likely a swan, cf. *canorus ales* 'the singing bird' in l. 15) (*Od*. II, 20, 10)
albus: album caprum 'a white goat' (*Od*. III, 8, 6)
albus: albus ... Iapyx 'the white Iapyx' (a favorable, or clean, cloudless wind, cf. *Od*. I, 3, 4) (*Od*. III, 27, 19–20). Cf.
albicant 'are, appear white': *nec prata canis albicant pruinis* 'and the fields are no longer white with hoarfrost' (*Od*. I, 4, 14)
ater: atra nubes 'a black cloud' (*Od*. II, 16, 2)
ater: atra nube 'with a black cloud' (*Od*. III, 16, 2)
ater: atra nocte 'in the black night' (*Epod*. 10, 9)
ater: atris viperis 'from black snakes' (*Od*. III, 29, 43–44)
ater: ater ... sinus Hadriae 'the black bosom [gulf] (of the Adriatic Sea)' (*Od*. III, 27, 18–19)
aureus: sidus aureum 'the golden star' (*Epod*. 17, 41)
candidus: nive candidum Soracte 'Mt. Soracte white with snow' (*Od*. I, 9, 1–2; Soracte is the famous mountain visible from Horace's villa near Tibur, the present Tivoli)
candidus: nive candidam Thracen 'Thracia white with snow' (*Od*. III, 25, 10)
candidis ... stellis 'with (your) white (i.e. clear) stars' (*Od*. III, 15, 6)
canus: canis ... pruinis 'with white hoarfrost [rime]' (*Od*. I, 4, 14)
flavus: flavum Tiberim 'the yellow [muddy?] Tiber' (the river that runs through Rome) (*Od*. I, 2, 13)
flavus: flavum Tiberim 'the yellow Tiber' (*Od*. I, 8, 8)
flavus: flavus ... Tiberis 'the yellow Tiber' (*Od*. II, 3, 18)
lividus: lividos ... racemos 'the bluish-black grapes' (*Od*. II, 5, 10–11)
niger: nigris ... ventis 'by the black winds' (*Od*. I, 5, 7)
niger: nigris silvis 'by the dark forests' (*Od*. I, 21, 7–8)
niger: nigri ... colles 'the black hills' (*Od*. IV, 11–12)
niger: nigro ... fumo 'from black smoke' (*Od*. III, 6, 4)
niger: nigro ... gregi 'to [his, i.e. Mercury's] black herd' (i.e. the netherworld population) (*Od*., 1, 24, 18)
niger: aequoris nigri 'of the black waters' (*Od*. III, 27, 23)
niger: nigro ... Orco 'the black Orcus [netherworld]' (*Od*. IV, 2, 23–24)
niger: nigrae ... frondis 'of its black [i.e. dark] foliage' (*Od*. IV, 4, 59)
niger: niger ... Eurus 'the black [i.e. not favorable] Eurus' (an East wind) (*Epod*. 10, 5)
niger: nigris ... barris '[eminently worthy of] dark elephants' (*Epod*. 12, 1)
niger: nigrum ... solem 'a black sun' (*Serm*. 9, 71)
pullus: *pulla ficus* 'the dark [i.e. ripe] fig' (*Epod*. 16, 46)
pullus: *hedera pulla* 'over the dark green ivy' (*Od*. I, 25, 18)
purpureus: mare purpureum 'the purple sea' (*Od*. II, 12, 2–3) (Cf. Homer's *oínopa pónton*; see also below, footnote 17)
purpureus: flos purpureus 'purple flower' (*Od*. III, 15, 15)

purpureus: purpureis . . . oloribus ‘with your [Venus’] purple swans’ (*Od*. IV, 1, 10)
ravus: rava lupa ‘a gray she-wolf’ (*Od*. III, 27, 3)
ravus: ravos . . . leones ‘the yellowish-gray lions’ (*Epod*. 12, 33)
rubens: Luna rubens ‘the red-shimmering Moon’ (*Od*. II, 11, 10)
ruber: Oceano rubro ‘the Red Sea’ (*Od*. I, 25, 32)
virens: hedera . . . virenti ‘by the green ivy’ (*Od*. I, 25, 17–18)
virens: virentis campos ‘the greening fields’ (*Od*. II, 5, 4–5)
virens: virens . . . flamma ‘the green [i.e. lively] fire’ (*Epod*. 17, 33)
viridis: viridi sub arbuto ‘under a green bush’ (*Od*. I, 1, 3)
viridis: viridis colubras ‘green [i.e. poisonous] snakes’ (*Od*. I, 17, 8)
viridis: viridi . . . myrto ‘with green [i.e. fresh] myrtle’ (*Od*. I, 4, 9)
viridis: viridis Cragi ‘of green Cragus [a mountain in Lycia]’ (*Od*. I, 21, 8)
viridis: viridi . . . Venafro ‘in green Venafrum [now Venafro, a town in Campania]’ (*Od*. II, 6, 15–16)
viridis: virides lacertae ‘the green lizards’ (*Od*. I, 23, 6–7)
viridis: viridi . . . pampino ‘with green [i.e. fresh] vine’ (*Od*. III, 25, 20)
viridis: viridi . . . pampino ‘with green [i.e. fresh] vine’ (*Od*. IV, 8, 33)
vitreus: vitreo . . . ponto ‘to the clear sea’ (*Od*. IV, 2, 3–4)

7.3.2 Body Parts & Bodies

albus: albo . . . umero ‘with her white shoulder(s)’ (*Od*. I, 5,19)
albus: bubus . . . albis ‘by white oxen’ (*Carm. Saec*. 49)
albus: capillus albus ‘(your) hair (is) white’ (*Epod*. 17, 18). Cf.
albescens: ‘whitening’ *albescens . . . capillus* ‘(one’s) hair turning white’ (*Od*. III, 14, 25)
ater: atro dente ‘with (a) black tooth/teeth’ (*Epod*. 6, 15)
ater: dens ater ‘a black tooth, black teeth’ (*Epod*. 8, 3)
ater: atro . . . cruore ‘by the black blood’ (*Epod*. 17, 31–32)
candens: candentis umeros ‘(your) white, i.e. splendid shoulders’ (*Od*. I, 2, 31)
candidus: candidos umeros ‘white shoulders’ (*Od*. I, 13, 9–10)
candidus: candidae cervici ‘on (your) white neck’ (*Od*. III, 9, 2–3)
candidus: puellae candidae ‘for a dazzling maiden’ (*Epod*. 11, 27)
canus: canos . . . capillos ‘white hair’ (*Od*. II, 11, 15)
cereus: bracchia cerea ‘wax-colored [i.e. nicely pale, white] arms’[9] (*Od*. I, 13, 2–3)
flavus: flavam comam ‘auburn hair’ (*Od*. I, 5, 4)
flavus: Phyllidis flavae ‘of blond Phyllis’ (*Od*. II, 4, 14)
flavus: flava Chloe ‘reddish-blond Chloe’ (*Od*. III, 9, 19)
flavus: in Ganymede flavo ‘in blond Ganymedes’ (*Od*. IV, 4, 4)

[9] Or ‘clear, smooth’ (Garrison 1991: 223). (The Romans did not appreciate a suntan, neither in men nor in women).

fulvus: niveus … cetera fulvus '(having a) white (mark on its forehead), the rest (being) reddish brown [said of a calf]' (*Od*. IV, 2, 59–60)
fulvus: fulvae matris 'of (its) brown-yellow mother [a lioness]' (*Od*. IV, 4, 14)
fulvus: fulvus Lacon 'the tawny Spartan' (a breed of sheep- or cattle-dog) (*Epod*. 6, 5)
lividus: livida … bracchia 'black-and-blue arms' (i.e. bruised from training) (*Od*. I, 8, 10–11)
lividus: dente livido 'with a (blue-)black tooth' (*Epod*. 5, 47)
luridus: luridi dentes '(your) yellowed teeth' (*Od*. IV, 13, 10–11)
luridus: pelle … lurida 'with yellow pelt' (*Epod*. 17, 22)
luteus: pallor luteus 'a yellowish pallor' (*Epod*. 10, 16)
murreus: murreum … crinem '(her) golden hair' (*Od*. III, 14, 22)
niger: nigris oculis nigroque crine 'with [his] black eyes and black hair' (*Od*. 1, 32, 10–11)
niger: dente … nigro 'with (your) black [i.e. ugly] tooth/teeth' (*Od*. II, 8, 3)
niveus: niveo colore 'by [her body's] snow-white color' (*Od*. II, 4, 3)
niveus: niveum … latus 'the white flank' (*Od*. III, 27, 25–26)
niveus: niveus … cetera fulvus '(having a) white (mark), for the rest being reddish brown [of a calf]' (*Od*. IV, 2, 59–60)
purpureus: purpureo ore 'with [his, i.e. Augustus'] mouth purple' (from wine?) (*Od*. III, 3, 12)
roseus: cervicem roseam '(his) rosy neck' (*Od*. I, 13, 2–3)
ruber: rubro sanguine 'with (its) red blood' (*Od*. III, 13, 6)
viola: tinctus viola pallor 'made up in pale yellow' [i.e. the color of the flower; cf. Garrison 1991: 311] (*Od*. III, 10, 14)
virens: virentis Chiae 'of the green [i.e. lush] woman from Chios' (*Od*. IV, 13, 6–7)
viridis: viridis … comas 'the green [i.e. watery] hair (of the Nereids)' (*Od*. III, 28, 10)

7.3.3 General Objects

albus: *albo panno* 'by a white cloth' (*Od*. I, 35, 21–22)
ater: atris ignibus 'with black flames' (*Epod*. 5, 82)
aureus: aureo cornu 'with [your, i.e. Bacchus'] golden horn' (*Od*. II, 19, 29–30)
aureus: virga … aurea 'with [your, i.e. Mercury's] golden scepter' (*Od*. I, 10, 18–19)
aureus: aureis … culullis 'from golden goblets' (*Od*. I, 31, 10–11)
eburnus: eburna … cum lyra 'with (your) ivory-colored lyre' (*Od*. II, 11, 22)
coruscus: igni corusco 'with red-flickering fire' (*Od*. I, 34, 6)
ruber: rubros … pannos 'the [blood-stained] red rags' (*Epod*. 17, 51)

7.3.4 Related Usage (Mostly Metaphorical)

albus: albus . . . pallor 'a white paleness' (*Epod*. 7, 15)
ardens: Vulcanus ardens 'red-hot Vulcan' (*Od*. I, 4, 8)
ater: morti atrae 'black Death' (*Od*. I, 27, 13)
ater: atrum venenum 'black [i.e. deadly] poison' (*Od*. I, 37, 28)
ater: fila . . . atra 'the black threads' (spun by Fate) (*Od*, II, 3, 16)
ater: atras . . . auris '(its) black [i.e. ugly] ears' (*Od*. II, 13, 34–35)
ater: ater . . . Cocytos 'the black Cocytos' [a netherworld river] (*Od*. II, 14, 17–18)
ater: atra Cura 'black Worry [the goddess]' (*Od*. III, 1, 40)
ater: atras curas 'black worries' (*Od*. III, 14, 13–14)
ater: atrae . . . curae 'black worries' (*Od*. IV, 11, 35–36)
aureus: te .. aurea 'you, the golden [maiden]' (*Od*. I, 5, 9)
aureus: auream . . . mediocritatem 'the golden middle road' (*Od*. II, 8, 5)
aureus: aureo . . . plectro 'with (your) golden [i.e. divinely inspired] plectrum' (*Od*. II, 13, 25–27)
aureus: tempus aureum 'the golden age' (*Epod*. 16, 64)
caeruleus: caerulea pube 'by the blue-eyed soldiery' (i.e. the Cimbri & Teutones, according to Garrison 1991) (*Epod*. 16, 9)
candidus: candidum . . . ducem 'a shining leader' (*Epod*. 3, 9–10)
candidus: candidum . . . ingenium 'the sincere mind' (*Epod*. 11, 1–12)
candidus: candide Maecenas 'my noble friend Maecenas' (*Epod*. 14, 5)
canities . . . morosa 'cranky white-hairedness' (i.e. 'old age', as opposed to *virens* for 'youth') (*Od*. I, 9, 17)
cressus: cressa nota 'a white [i.e. festive] mark' (on the calendar) (*Od*. I, 35, 10)
furvus: furvae . . . Proserpinae 'of somber Proserpina' (*Od*. II, 13, 21)
igneus: igneam . . . aestatem 'the red-hot summer' (*Od*. I, 17, 3–4)
lividus: lividas . . . obliviones 'black (forgotten) memories' (*Od*. IV, 9, 33–34)
niger: nigrum Merionem 'the black [i.e. infamous] Meriones' (the companion and later henchman of Idomeneus, the Cretan king) (*Od*. I, 6, 15)
niger: nigrorum . . . ignium 'of the dark fires' (*Od*. IV, 12, 26)
nitens: nitentis . . . Cycladas 'the dazzling white Cyclades' (*Od*. I, 14, 19–20)
nitidus: nitido . . . adultero 'the dazzling adulterer' (*Od*. III, 24, 20)
nitidus: nitido curru 'in (the Sun's) splendid course' (*Carm. Saec*. 9)
nitor: Glycerae nitor splendentis 'splendid Glycera's dazzling whiteness [i.e. beauty]' (*Od*. I, 19, 5–6)
pallidus: pallida mors 'pale Death' (I, 4, 13)
purpureus: purpureo colore '[Autumn] with (its) purple color' (*Od*. II, 5, 12)
rubens: rubente dextera 'with [his, i.e. Juppiter's] red [i.e. fiery] right hand' (*Od*. I, 2, 2–3) Cf.
non erubescendis ignibus 'because of fires that should not make [you] turn red' (i.e. a passion that you need not be ashamed of) (*Od*. I, 27, 15–16)
virens: donec virenti canities abest 'as long as (you, the) greening [i.e. young] (scil. Thaliarchus) can keep old age away' (*Od*. I, 9, 17)
vitreus: vitream Circen 'the [crystalline, i.e.] splendid Circe' (*Od*. I, 17, 20)

7.3.5 Unspecified

mare ... decoloravere caedes 'the slaughter has changed the color of the sea' (*Od*. II, 1, 35)
nullus color '(buried money) has no color' (*Od*. II, 2, 1)
amissos colores 'its [the wool's] lost colors' (*Od*. III, 5, 27)
liquidis ... coloribus '[he [scil. Parrhasius] painted] with wet colors' (*Od*. IV, 8, 7)
color ... rosae 'the color of the rose' (*Od*. IV, 10, 4) (Cf. *roseus*)
color stercore fucatus '(crocodile-)manure applied color' (*Epod*. 12, 10–11)

8 Analyzing the Distribution

8.1 Preliminaries

Looking at the words for colors that we find displayed in Horace's verse, it strikes us immediately that the overwhelming majority of color terms have to do with various expressions for 'black' and 'white'. As to 'black', the front runner is *ater*, with 17 out of 139 total tabulated occurrences; runner up is *niger*, with 16 occurrences. If we consider *lividus* '(blue-)black' as a legitimate representative of this color,[10] we can add its 4 to the total to obtain 37 tokens of terms for 'black'. Adding the two occurrences of *pullus* 'dark' gives us a total of 39 out of 139, or 27.8 %.

As to Horace's expressions for 'white', we find a 'dazzling' array of terms, eight altogether.[11] The real color words are best represented: the most frequently used term is *albus*, with 11 occurrences (13 with its derivates such as *albescere, albicare*), but *candidus* is very frequent as well: 9 occurrences (10 including *candens*); the adjective is also used metaphorically (meaning 'splendid', 'honest', and so on). Quite popular are *canus* (with *canities*) and *nitidus* (with related items), with 3, respectively 4 occurrences. The next in line, *niveus*, has 3 occurrences, while *pallidus* (including *pallor*) shows 2. A few single occurrences such as *cereus, cressus*, *eburnus* bring up the rear. Altogether the words for 'white' and its cognate color shades account for 38 out of the total of 139 occurrences of color terms, or 27.4 %.

The third most frequent color to be cited is 'green', mainly represented by the adjectives *viridis* and the 'defunct' participle *virens* (lit. 'greening'; often used

[10] *ater* and *lividus* carry additional connotations such as 'doom' and 'malice'.

[11] The terms for 'dazzling', such as *splendidus, splendens, fulgens,* etc. have not been included in the spectrum for 'white', although one could argue that they might have their proper place here, especially in view of the contrasts that Horace makes use of in relation to other colors. (Cf. *Od*. III, 13, where the glittering splendor of Horace's favorite spring of unknown location (*fons Bandusiae, splendidior vitro* 'the Bandusian spring, clearer than glass') contrasts with the red blood (*rubro sanguine*) of the animal to be sacrificed to/in it). See also below, Sect. 11.

metaphorically as a synonym for 'young'). Altogether the 14 occurrences of words for 'green' make up 9.9 % of the total color vocabulary.

Among the remainder of the color terms, we find a few, more unusual color words; many of these are single occurrences, so-called *hapax legomena*. Most conspicuous in this connection is the unique occurrence of the word for 'blue', *caeruleus*, on which I will have more to say below, in Sect. 10. A number of terms for 'mixed' colors are also encountered, most of them belonging in the yellow-brown and/or darker bands of the spectrum. Thus, we have *fulvus* 'brownish-yellow', *furvus* 'dark, brown', also used metaphorically: 'ominous' (see below), *luteus* 'yellow(ish)', *luridus* 'yellow(ed)', also used metaphorically: 'sinister', *murreus* 'golden brown', *ravus* 'yellowish gray', and so on. Note also the remarkable occurrence of the substantive *stercus* 'manure', used to signify a brownish color, as in *color stercore fucatus* 'a color resulting from the application of (crocodile) manure'.[12]

Unusually few occurrences (from a modern perspective) are found of terms for 'red': we find *ruber* only three times (together with the derivatives *rubens* 'reddish', *erubescere* 'become red [e.g. in the face]', we reach a total of 6). To this, add a single occurrence of *roseus* 'rosy, pink' and the mention of *color rosae*, 'the color of the rose', once. I will come back to this phenomenon below, Sect. 11

A number of terms are based on derivation from substantives, such as *aureus* 'golden' (8 times), *cereus* 'wax-colored' (once), *igneus* 'fiery' (once), *purpureus* 'purple' (4 times), *vitreus* 'clear [like glass], splendid' (twice), and so on.

Sometimes Horace uses a paraphrase, perhaps in an effort to capture an unusual or 'loaded' color; an example is found in the expression from the *Epodes* (12, 10–11): *color stercore fucatus* (see above). Another is the paraphrase of a color ('yellowish'),[13] apparently derived from *viola*, the violet; here, a somewhat more unusual flower ('violet') than the ubiquitous *rosa* ('rose') has been employed for greater effect: *tinctus viola pallor* '[the lovers'] pale complexion, due to an application of [an extract of] violet' (*Od.* III, 10, 14).

8.2 A World in Black and White

Given the preponderance of the terms for 'black' and 'white', I will, below, first list the occurrences of these words according to their major collocations. Here, I will use the groups that I established in Sect. 7.3, above. Table 1 shows the use the terms for 'black'.

As a result, we can find 33 instances of 'black'. Excluding the category Metaphor, the number is 23.

As regards the terms for 'white', Table 2 shows the uses in the corpus.

---

[12] Norwegian has similarly a commonly recognized (though highly restricted) term for this color: *bæsjebrunt,* literally 'shit-brown' (often used in older farm houses for painting floors).

[13] Thus according to Garrison (1991); for different opinions, see Clarke (2003: 123, 170).

Table 1 Appearance of the terms for 'black' in the corpus

	Nature	Body	Object	Metaphor
Ater	5	3	1	8
Niger	11	3	0	2
Total	***16***	***6***	***1***	***10***

Table 2 Appearance of the terms for 'white' in the corpus

	Nature	Body	Object	Metaphor
Albus	7	4	1	1
Candidus	3	4	0	3
Canus	1	1	0	1
Nitidus	0	0	0	4
Niveus	0	0	3	0
Pallidus	0	0	1	1
Cereus	0	0	1	0
Cressus	0	0	0	1
Eburnus	0	0	1	0
Total	***11***	***9***	***7***	***11***

The results show that 'white' has 38 instances, 27 if the category metaphor is excluded.

Adding up the figures for black and white we obtain the following results: the instances of color terms are 33 for black and 38 for white.

If we cluster the three first categories (Nature, Body & Object) into one new category that we may call 'descriptive', the result is 23 occurrences for black and 27 for white.

On a total count of 139 color terms in my corpus, this means that the colors black and white together are represented by 71 instances out of 139, i.e. 53.2 % of the total number. The next most frequently occurring color term is that for 'green', *viridis* (and related terms) with only 14 occurrences (of which 4 are non-descriptive or metaphorical), which then accounts for only 9.9 % of the total number of color terms.

Clearly, the black and white colors (especially if we add their shades and nuances, such as *pullus* for 'dark green, black' or *nitidus* for white)[14] account for the vast majority of colors used in our corpus, both descriptively and metaphorically. The next step is how to interpret these findings in light of our initial question.

9 Was Horace Color-Blind?

With the reservations given above concerning the (cross-)cultural significance, both of terms for colors and of terms related to colors (as in 'color blind'), there seems to be no reasonable doubt that Horace, in his depiction of nature and people, *really* saw the colors he was referring to when he employed either black or white, or both.

[14] *furvus* 'dark, somber' could perhaps have been counted here as well; for 'white', one could have included such terms as *splendidus* or *splendens* 'splendid'.

Likewise, as far as the other colors are concerned, we may safely assume that what he wrote was what he saw (or occasionally, wanted to see); the problem is that (with the exception of *viridis* 'green') there aren't that many other color mentions in the first place, and that many of them are being used rather conventionally—most often, metaphorically.

In fact, some of the most beautiful descriptive passages in the *Odes* make reference precisely to either or both of the colors black and white (cf. the already quoted *Vides ut alta stet nive candidum/Soracte*, from *Od*. I, 9, 1–2, or the equally beloved, elegant Spring Ode from Book I, *Solvitur acris hiems* . . . ('the bitter winter is dissolved. . .'), where the poet describes the post-wintery 'fields as no longer white with rime': *nec prata canis albicant pruinis*; *Od*. I, 4, 4). And when Horace ridicules former flames, now saddled with the weaknesses of age, we have little trouble seeing Lyce's, Barine's, and other former mistresses' bad teeth being 'adorned' with the epithet 'black' (either *niger* as in *Od*. II, 8, 3, or *ater*, as in *Epod*. 6, 15 and 8, 3, or *lividus*, as in *Epod*. 5, 47).[15] Similarly, his characterization of the Adriatic Sea, that common graveyard of Roman sailors, especially during the raging storms of autumn (cf. *Od*. II, 14, 14), as *ater sinus Hadriae* 'the black bosom of the Adria' (*Od*. III, 27, 18–19) seems particularly appropriate.

When it comes to the other colors, however, the poet seems not always capable of painting a convincing picture. As for the occurrences of 'green' (*viridis, virens*), the next most frequent item on our list, they are mostly standard. We don't have to use much of our imaginative force to represent fields or bushes or even lizards (*virides lacertae*, *Od*. I, 23, 6–7) as 'green'; plants such as myrtle, ivy and the ubiquitous vine are (ever-)green by nature (cf. *Od*. I, 4, 9; I, 25, 17–18; II, 25, 20; IV, 8, 33; and *passim*). In addition, many of Horace's 'greens' are clearly metaphorical (denoting 'youth', mostly) and carry little descriptive value (in the sense defined above).

Moreover, among the adjectives that are used to represent what one could call a 'true' color, we find quite a few that are decidedly conventional, even to the degree that they represent a fixed, 'locked' combination Adj + N, where the adjective functions as the classical trope called *epitheton ornans*, popular ever since Homer. Thus, we find regular references to the 'yellow(ish) Tiber' (*flavus Tiberis*, e.g. *Od*. I, 2, 13; I, 8, 8; II, 3, 18)—an epithet that certainly makes sense if one thinks of the Tiber's (even today) muddy waters, and one which had been the river's standard attribute ever since the days of Ennius (from whom Horace probably had borrowed the expression).[16] Similarly, we encounter a mention of the sea as being 'purple': one is immediately reminded of Homer's standard expression *oînops*

[15] As Nesbit & Hubbard remark, for Horace bad teeth are the "hallmark of decaying courtesans". (1978: 125).

[16] I am indebted to Steven J. Willett, Shizuoka University, Japan, for this observation.

Interestingly, the river has retained its magic, both color- and otherwise, all the way into our own times. Compare how Canadian novelist Margaret Atwood, in her novel *The Blind Assassin* (2000), has her heroine, while honeymooning in Rome, look down from her hotel window on "the Tiber, floating along, yellow as jaundice" (p. 304).

póntos ‘the wine-colored sea’ (ε 132 *et passim*). There, the occurrence of the qualifying adjective is as familiar and expected as is Horace’s use of *clarus* ‘clear, brilliant’ in e.g. *claram Rhodon aut Mytilenen* (about two Greek islands in the Ionian sea; *Od.* I, 7, 1)—itself a direct replica of Homer’s expression *Lakedaímona dîan* ‘splendid Lacedaemon [=Sparta]’ (δ 702, ε 20 *et passim*).[17]

Thus, *virens* ‘green, vigorous’ in *donec virenti canities abest* ‘as long as [your] youth lacks the sorrowful grey mane’ (*Od.* I, 9, 17) or in *virentis Chiae* ‘the lush woman from Chios’ (*Od.* IV, 13, 6–7) have very little to do with the color ‘green’ as such. And on the reportedly ‘green hair’ of the water nymphs, the Nereids (*viridis* . . . *comas*; *Od.* III, 28, 10), the jury is still out: maybe their hairdos are thought to be ‘green by association’ (with the ‘greenish’ water)?

There are only a few places where the colors outside the white-black gamut seem to have been used with some precision. One instance is the occurrence of *lividus*, when Horace identifies the ‘blue-black grapes’ (*lividos* . . . *racemos*, *Od.* II, 5, 10–11), or where a fig is qualified as *pulla* ‘dark [green]’ (*Epod.* 16, 46)—this adjective, too, properly belongs in the ‘dark’ part of Horace’s color spectrum, as we have seen earlier. In a number of cases, an unexpected color turns out to be less than descriptive-pictorial, e.g. when Venus’ ‘purple swans’ (*Od.* IV, 1, 10: *purpureis oloribus*) turn out to be more ‘splendid’ than really ‘purple’.[18]

10 How Blue the Sky. . .

One of the more astonishing features of Horace’s color usage is the absence of references to the sky as blue. This may be a common characteristic of Roman poetry (cf. Clarke 2003: 47ff. on *caeruleus* ‘blue’ (from the word for ‘sky’, *caelum*)), but the fact remains that the sky’s blue brightness does not seem to have evoked any poetic resonance in our poet. One banal explanation could be that the Romans didn’t have any chance to see the blue sky, due to their persistent use of firewood for cooking and heating, which could have resulted in a massive brownish haze over the city (akin to what one sees when approaching a modern Far Eastern city such as Bangkok or Beijing from the air). But then again, Horace did not suffer from such restrictions in his country hideaway, and he was certainly able to see his beloved Soracte, unimpeded by the smoke clouds of the not too far away *Vrbs*.

Another possible explanation would draw attention to the fact that the sky’s blue color was not thought of as a suitable expression for feelings and mental imagery among the Romans, who often used Nature in their poetry to reflect their own states

[17] On Homer’s ‘wine-dark’ sea and its many (mis-)interpretations, see Maxwell-Stuart (1981), esp. pp. 6ff.

[18] Clarke mentions that many scholars have ‘downplayed’ the value of this color word and instead, suggest that we translate *purpureus* as ‘gleaming’ or ‘silver’ (swans). Clarke herself does not buy into this, however (2003: 292).

of mind—which would not, as we have seen, put them under any strict obligation to render the true colors of natural objects or humans faithfully and/or completely.

Also, if the sky happens to be blue, as it is usually the case (with the proviso mentioned above) in Mediterranean countries, one perhaps doesn't pay too much attention to this fact. The almost idolatrous veneration for blue skies and golden sunlight that characterizes Nordic people is totally absent from the mentality of those whose relationship to sun and skies is of a more restrictive and restricted character. To imagine the modern activity of sunbathing as occurring in a classical environment seems as ridiculous as it is anachronistic, and we have to travel hundreds of centuries (in fact all the way to Goethe) to find a mental link between the sun's appearance in the blue skies and one's (re-emerging) belief in God ("die Sonne scheint heiß, und man glaubt wieder einmal an einen Gott," as Goethe wrote down in his *Italienische Reise,* upon descending, on September 11, 1786, from the foggy and icy Brenner Pass into the Alto Adige valley and the plains around Bolzano).

Yet, one wonders. The only occurrence of the word for 'blue', *caeruleus*, in my Horace-corpus is that in *Epod*. 16,9, where Horace talks about the *caerulea pubes*, 'the blue-eyed soldiery', apparently referring to the Germanic tribes that nearly overran Rome in the 2d century B.C.[19] Moreover, not even the sea was blue for Horace, as far as we know; the epithets he uses ('black', 'purple', 'shining', 'red') are more in line with the established patterns originally due to Homer and later Virgil, as we have seen in the examples above (the 'purple sea', the 'black Adria', and so on). And of course, Horace would not have been personally in contact with those 'blue-eyed' Germans—he is merely referring to a long-standing, historic-literary tradition.

11 Other Colors

The absence of one of our primary colors from Horace's nature palette is all the more telling, given that another color, red, is also represented very sparsely. Actually, the only time we really 'see red' in his poems is when he mentions his beloved *fons Bandusiae*, the 'Bandusian wellspring' he apostrophizes as being of 'more than crystal-like splendor' (*splendidior vitro*; *Od*. III, 13, 1), and to which he promises to sacrifice a young billy-goat the next day: *cras donaberis haedo* 'tomorrow you will be given a kid' (*Od*. III, 13, 3). It is this animal's red blood that will adorn the spring's clear waters (*inficiet tibi/rubro sanguine rivos* '[the kid's] red blood will color your banks'; *Od*. III, 13, 6–7).

The other words having to do with the color 'red' are all of a less distinct character: the moon is called *rubens* 'red-shimmering' (*Od*. II, 11, 10), but so is

[19] While Clarke does not mention this occurrence in her 2003 conspectus of color terms (where she lists only examples from the *Odes*), she does mention *caeruleus*/*caerulus* as being derived from Catullus and Propertius (pp. 47–49).

Garrison (1991) specifically refers to the invasion of the Cimbri & Teutones in 101 B.C.

Jupiter's right hand, said to throw fiery bolts of lightning on the City: *rubente dextera* 'with his flaming right hand' (*Od*. I, 2, 2–3). In another context, we find a certainly metaphorical, maybe even metonymical, derived use of 'red' as the color of shame (*non erubescendis ignibus* 'by the flames [of a passion] that you shouldn't have to be ashamed of [because the object of your desire is a slave girl]') (*Od*. I, 27, 15–16). Vulcan is said to be *ardens* ('burning, red-hot') when he 'stokes up his workshops': *Vulcanus ardens/urit officinas* (*Od*. 1, 4, 8); elsewhere the summer is referred to as *igneus* 'red-hot, fiery' (*Od*. I, 17, 3–4).

Apart from the case of the sacrificial kid in *Od*. III, 13, mentioned above, the only other 'objectively' descriptive use of the word for 'red' occurs in the 17th *Epode*, where Horace ironically addresses a reluctant mistress, Canidia, by referring to her jumping out of childbed in pursuit of more interesting exploits, even 'while the midwife is busy washing your blood-stained sheets' (*... tuo/cruore rubros obstetrix pannos lavit, ...*; *Epod*. 17, 51). And the use of *ruber* in *Oceano rubro* 'the red Ocean' (*Od*. I, 25, 32) may be better explained, in its context of geographical locations, as a fixed expression, cf. our 'Red Sea'.[20] Finally, the derived word *roseus* 'rosy, pink' (from *rosa*) is neither color-specific (the rose is also said to be a 'purple flower', *flos purpureus*), nor other than conventionally descriptive, as in *color rosae* 'the color of the rose' (*Od*. IV, 10, 4); cf. the mention of Telephus' 'rosy neck', *cervicem roseam* (*Od*. I, 13, 2; incidentally, a *hapax* in my corpus), which is every bit as conventional as the poet's later reference to Lydia's 'splendid neck': *candidae cervici* (in *Od*. III, 9, 2–3).

Many of the other color terms that we find in Horace have too limited,[21] or too conventional a distribution to be of much use, when evaluating the poet's 'feel' for nuances of color. As an instance, take the word *fulvus*, 'brownish-yellow', traditionally used to denote the color of a lion's (occasionally a dog's) pelt. True enough, Horace uses the term three times for animals: a lioness, *fulvae matris* 'of the brownish mother' (*Od*. IV, 4, 14), a calf, said to be brown except for a white mark: *niveus videri, cetera fulvus* 'showing white [in one place], but brown otherwise' (*Od*. IV, 2, 59–60), and a Spartan cattle dog, *fulvus Lacon* 'a brown Laconian' (*Epod*. 6, 5).[22] In line with other color *hapaxes*, however, these usages do not reveal much about the way these colors were attributed 'normally' (i.e. outside their fixed environments).[23]

[20] As does Clarke (2003: 140).

[21] As mentioned earlier, a number of them are *hapax legomena*.

[22] Clarke (2003: 91) quotes an earlier mention by André (1949: 133), who asserts that the color *fulvus* is traditionally ascribed to lions, especially in Virgil (see also Clarke 2003: 165).

[23] Examples include *murreus, luteus, furvus, cressus* and others. For instances of their usage, see the listing in Section 7.3.

12 Conclusion

Having considered the varying usage that Horace makes of the available color terms in his language, two thoughts present themselves as worthy to take home and keep for further reflection.

One concerns the fact that there is indeed a great numerical disparity between the terms being used for 'regular' colors and those from the predominantly black/white scale. Over one half of all the color terms found fall into this latter category, whereas other color terms (among them some very common ones, such as blue and red) are represented only once or sporadically; while still others, among these many for 'blurred' colors such as yellowish-red, grayish-yellow, and so on, occur mostly in certain isolated, predictable environments (often as *epitheta ornantia*).

On the other hand, the 'true-to-life' depiction of Nature with which we have become familiar ever since Antiquity, and which certainly was appreciated among the Greeks and Romans as far as painting goes, does not seem to have triggered a similar need when it comes to writing and 'pictorial', literary description. Many of the color words I have registered (over one third of the total occurrences) are used in a metaphorical sense; and among the ones used for description, as we have seen, not many correspond to what we see as 'true' colors. For instance, we wouldn't learn, from reading Horace, whether the skies were actually perceived as blue by the Romans.[24]

Concluding, then, I cannot prove, or even reasonably maintain, that Horace was color-blind in the strict sense of the word.[25]

What we do notice is that Horace's use of the color spectrum is markedly different from ours; the difference manifests itself especially in the descriptive aspects of his poetry, where black and white dominate. So, perhaps one could venture the conclusion that Horace's 'color-blindness' was, at least partially, culturally conditioned and could be due to a "different appreciation" of the color schema, as the late University of Texas classicist Alice Kober (1934: 189) has suggested.

In addition, I would venture the hypothesis that the poet's treatment of the other colors lacks in variety because he seems to have been perfectly contented with the effects he could obtain by 'painting' in black and white. After all, even in our own days, the effects realized in older 'B&W' movies often widely surpass the impact of more modern techniques, merely serving an intensified chase for a 'truer', life-like coloring. Here, it behooves us to recall that even a successful color chase does not guarantee an artistically satisfying catch; a motif's beauty is not automatically enhanced by objects being painstakingly depicted in all minute details. Horace

[24] Some other Roman poets do have a few references to this specific color, when it comes to the heavens above; 'blue' is also sometimes used for the color of the sea (cf. Clarke 2003: 47).

[25] Neither were, presumably, the people of antiquity in general, as among others, William Gladstone has maintained (Gladstone 1877; I owe this reference to Jacqueline Clarke, in personal communication).

Gladstone's hypothesis of a "cultural deficiency" among the Greeks and Romans (not to speak of passing references to those early Indo-Europeans still being in their cultural 'infancies') has been resoundingly refuted by scholars such as Irwin (1974: 201–203) and Maxwell-Stuart (1981).

paints with a light stroke, giving us an outline rather than a blueprint. Situations, happenings, and relationships are visualized in tones of light and dark, in shades of density, rather than in a panoply of assorted colors. The things that remain unsaid, such as the 'missing' colors, are left for us, the readers, to fill in for ourselves. This opening towards a "readerly cooperation" (Mey 2000: 148ff) is among the features that make Horace's work unique, and truly *aere perennius* 'more lasting than bronze' (*Od*. II, 30, 1).

Acknowledgements An earlier, much shorter version of this paper appeared as Mey 2005a; a revised version as Mey 2005b. For the present, thoroughly revised and enlarged version, I owe thanks to Professor Jacqueline Clarke of the University of Adelaide, South Australia, who provided a number of useful comments and corrections.

References

André, J. (1949). *Etude sur les termes de couleur dans la langue latine*. Paris: Klincksieck.

Atwood, M. (2000). *The blind assassin*. New York: Anchor Books.

Berlin, B., & Kay, P. (1991 [1969]). *Basic color terms: Their universality and evolution*. Stanford: CSLI Publications.

Clarke, J. (2003). *Imagery of colour & shining in Catullus, Propertius, & Horace* (Lang classical studies, Vol. 13). New York: Peter Lang.

Edgeworth, R. J. (1992). *The colors of the Aeneid*. New York: Peter Lang.

Garrison, D. H. (1991). *Horace: Epodes and Odes. A new annotated* (Latin ed.). Norman/London: University of Oklahoma Press.

Gladstone, W. E. (1877). The colour sense. *The Nineteenth Century, 2*, 366–388.

Hjelmslev, L. (1943 [1993]). *Omkring sprogteoriens grundlæggelse*. Copenhagen: Munksgaard/Nordisk Sprog- og Kulturforlag. (English translation by Francis J. Whitfield as *Prolegomena to a theory of language*. Madison: University of Wisconsin Press, 1953)

Horace [Q. Horatius Flaccus]. (1955). *Opera* (E. C. Wickham, Ed.). Oxford: Clarendon Press.

Irwin, (M.) Eleanor. (1974). C*olour terms in Greek poetry*. Toronto: Hakkert.

Johnson, H. (2009). *Hugh Johnson's pocket wine book 2009*. London: Mitchell Beazley.

Jucker, A., & Taavitsainen, I. (2010). *Historical pragmatics*. Berlin: Mouton de Gruyter.

Kober, A. E. (1934). Some remarks on colour in Greek poetry. *Classical World, 27*, 189–191.

Maxwell-Stuart, P. G. (1981). *Studies in Greek colour terminology* (Glaukos, Vol. 1). Leiden: Brill.

Mey, J. L. (2000). *When voices clash: A study in literary pragmatics*. Berlin/New York: Mouton de Gruyter.

Mey, J. L. (2005a). *Candidum Soracte*: Was Horace color-blind? *RASK: International Journal of Language and Communication, 22*, 11–36.

Mey, J. L. (2005b). Horace and colors: A world in black and white. In D. Haug & E. Weilo (Eds.), *Haptacahaptaitis: Festschrift for Fridrik Thordarson on the occasion of his 77th birthday* (pp. 163–175). Oslo: Novus forlag & Instituttet for sammenlignende kulturforskning.

Nisbet, R. G. M., & Hubbard, M. (1970). *A commentary on Horace: Odes book I*. Oxford: Oxford University Press.

Nisbet, R. G. M., & Hubbard, M. (1978). *A commentary on Horace: Odes book II*. Oxford: Oxford University Press.

O'Neill, P. G. (1982). *Essential* ***kanji***. Tokyo/New York: Weatherhill.

Pullum, G. K. (1992). *The great Eskimo vocabulary Hoax and other irreverent essays on the study of language*. Chicago/London: University of Chicago Press.

Steinmayer, O., et al. (2000). *A web discussion of color terms in Latin*. http://omega.cohums.ohio-state.edu/mailing_lists/CLA-L/2000/02/0071.php

Tolstoy, L. N. (1952). *Anna Karenina*. Moskva: Izdatel'stvo Pravda. Tom I.

Self-Conscious Emotions in Collectivistic and Individualistic Cultures: A Contrastive Linguistic Perspective

Barbara Lewandowska-Tomaszczyk and Paul A. Wilson

Abstract The present paper focuses on linguistic and culture-bound aspects of the properties of *individualism* and *collectivism* through an English-Polish analysis of the emotions shame and guilt. Combining theoretical analyses with the analysis of authentic data, this work breaks new ground in cognitive-based language analysis in its pragmatic setting and attempts to shed new light on complex issues pertaining to cultural identities. The study presents an investigation on language corpus materials of English and Polish and furthermore it enriches the methodology with questionnaire-based (GRID) data of English and Polish, identifying cross-linguistic similarities and differences between the relevant dimensions and components with respect to shame and guilt. The corpus data used in previous studies show a stronger emphasis on self-construal at the individual level of identity with the Polish users, while the English users were presented to attend to a larger extent to the relational self derived from the interactional relations with others . The present study provides strong additional support for a more refined model of both collectivism and individualism and further elaborates on the assumptions of a contrastive analysis of self-conscious emotions.

Keywords Collectivism • Cross-linguistic analysis • English • Culture • *guilt* • Individualism • Polish • *shame*

B. Lewandowska-Tomaszczyk (✉) • P.A. Wilson
Institute of English Studies, University of Lodz, Kosciuszki 65, 90 514 Lodz, Poland
e-mail: blt@uni.lodz.pl

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2,
DOI 10.1007/978-3-319-06007-1_7, © Springer International Publishing Switzerland 2014

1 Contrastive Cognitive Corpus Linguistic Studies

The present paper focuses on the ways emotions are expressed in cross-linguistic and cross-disciplinary perspectives. More precisely, it is a contrastive corpus-based study of Polish and English emotion concepts and the linguistic patterns they enter to illuminate language and culture-bound aspects of the properties of *individualism* and *collectivism*. Combining theoretical analyses with the analysis of authentic data, this work breaks new ground in cognitive-based language analysis in its pragmatic setting and attempts to shed new light on complex issues pertaining to cultural identities.

The first attempt to look at English and Polish differences between cultural and national identity profiles was carried out with reference to two online English and Polish discussions of sports events (Lewandowska-Tomaszczyk in press), in which details of the dynamics of the encounters and the interactional effects they produce by the use of *language*- and *culture-specific* strategies were discussed and interpreted. The individualism-collectivism dimensions referred to turned out to be more varied and complex than originally postulated. The corpus data used in that study show a stronger emphasis on self-construal at the individual level of identity with the Polish users, while the English users were presented to attend to a larger extent to the relational self derived from interactional relations with others. Additional support for a more refined model of both collectivism and individualism can be found in a recent study on values and worldviews in the context of European Union countries.[1] The present study identifies cross-linguistic similarities and differences between the relevant dimensions and components with respect to shame and guilt in an investigation that is extended to larger language corpus materials of English and Polish and furthermore methodologically enriched with English and Polish data obtained from the GRID instrument (Fontaine et al. 2013).

2 Pragmatics and Emotions

Emotions are one of the most frequently investigated topics at present in the context of cognitive pragmatics[2] (Tracy et al. 2007). Michael Gilbert (1997) proposes four foundations of the pragma-dialectic approach to emotions. They are externalization, functionalization, socialization, and dialectification. Each of these properties or principles relates to our idea of a (prototypical) Emotion Event as presented and discussed in Lewandowska-Tomaszczyk and Wilson (2013), albeit our model aims to represent a temporal sequence of event parameters and their causal relations, while what Gilbert discusses are aspects of the event not necessarily interpreted in terms of

[1] The survey was conducted by the Communication Department of Fundación BBVA and is accessible at www.fbbva.es

[2] Contextual accounts of actions involving emotion language can be identified both in social psychology literature as well as in pragmatic and cognitive linguistic approaches, starting with Fillmore's frames-and-scenes semantics (1977) to Sperber and Wilson (1995), Kövecses (2000), and Tracy et al. (2007), who deal with contextual aspects of emotion language.

the causal or temporal links. Gilbert's *Externalization* foundation refers to the expression of *meaning* by means of linguistic and paralinguistic signals we refer to as *exbodiment*. The parameter of *socialization* assumes an interaction between communicating individuals. In the scenario of a prototypical Emotion Event, this aspect will be primarily present in the case of social emotions, whose origin is connected with the situational or contextual factors involving an interactant, some of which will constitute a *stimulus for* or a *cause* of an emotion event. Basic emotions, such as fear, surprise or disgust, irrespective of the fact as to whether they are verbally and/or paralinguistically expressed or not, do not necessarily assume the presence of an interlocutor. However, self-conscious emotions, including shame and guilt, which are the topic of the present paper, presuppose the explicit or implicit presence of an interlocutor, although the degree of their socialization function does not need to be identical. While shame, for instance, is closely connected with the social other, guilt is less explicitly conditioned by such a real or imagined presence.

Functionalisation, the next of the four properties, refers to the presence of a 'purpose' of a communicative act. "Emotions – as proposed by Gilbert (1997) – serve a crucial purpose, they serve to communicate to one's dispute partner the degree of commitment, concern, and feeling one has about a given standpoint." The concept of a communicative purpose has to be juxtaposed with that of 'control'. Emotions are subject to conscious control to different degrees. Control is both speaker-bound to a certain extent as well as determined by a given culture and language. For that reason, the 'purpose' of showing an emotion can be interpreted only in relation to the element, which is its cause and stimulus, and which constitutes a causal link between the internal feeling of an emotion and its verbal expressivity. Even verbally expressed emotions are not always the subject of the experiencer's conscious control, so they have to be treated as having an inherent 'purpose' only in the sense of alerting the interlocutor's state of mind or state of action.

Dialectification, and its manifestation, *emotionalisation*, are – according to Gilbert – properties that make it possible to determine the conversational level of the speaker's *consistency* and *commitment to their intent*. Consistency is defined as a positive correlation between what is said and what is manifested non-verbally, while commitment is the degree of conviction the speaker has about the point s/he is making.

3 Self-Conscious Emotions Examined

Besides being interactional and social in nature, shame and guilt can be considered *moral emotions*. Highlighting both the social and moral features of these emotions, Beer (2007) observes that "in order to experience a self-conscious emotion, one must have an awareness of self (self-perception), an awareness that others are judging that self (person inference), and an awareness that there are a set of rules or social norms that determine whether the actions of the self are "right" or "wrong" (social norms)" (p. 53).

Moral emotions such as shame and guilt are primarily self-conscious emotions which deal with self-evaluative judgments, and posit us against others. It is for this

reason that we follow the convention set by others (e.g., Tracy et al. (2007) and Lewis (2008)) and refer to shame and guilt as self-conscious emotions.

Recent research (e.g., Tracy and Robins (2004)) has strived to address the imbalance in our understanding of self-conscious emotions (e.g., shame, guilt, pride and embarrassment) in comparison with the relatively greater knowledge we have gained on basic emotions – happiness, surprise, sadness, anger, fear and disgust. Learning about the nature of different classes of emotions, such as self-conscious emotions, is highlighted by Levenson's (1999) call for a number of emotion theories each pertaining to a specific category of emotions rather than one, overriding theory of emotion. In comparison with the biological basis (Shaver et al. 1987) of basic emotions, self-conscious emotions involve a greater degree of complex cognition (Tracy and Robins 2004). Other features that distinguish self-conscious emotions from basic emotions include the focus on the attainment of social goals (Keltner and Buswell 1997); their emergence relatively later in childhood (Izard et al. 1995; Lewis 1995); the involvement of self-evaluative processes (Buss 2001; Lewis et al. 1989); the lack of discrete facial expressions that are universally recognised (Tracy and Robins 2004); and relatively weaker universality (Tracy and Robins 2004). It is the latter quality of self-conscious emotions that is most relevant to our aim of using the GRID instrument and the cognitive corpus methodology in a contrastive analysis of the self-conscious emotions shame and guilt in Polish and British English. In the sections to follow, a discussion of individualistic and collectivistic cultures is first presented, followed by an account of shame and guilt from a cross-linguistic and cross-cultural perspective.

4 Individualism and Collectivism

4.1 Individualism

In individualistic cultures one perceives oneself as an individual, autonomous entity and there is less emphasis placed on one's relationships to others. The various accounts of individualism share the fundamental features of more of an individualised construal of goals, uniqueness and control (Oyserman et al. 2002). Highlighting the personal autonomy associated with individualism, Hofstede (1980) views the inclusion of self-fulfilment and personal accomplishments in one's identity, the importance of rights in comparison with duties, and a focus on oneself and immediate family as central features. In contrast with collectivistic individuals who have relatively more interdependence within their in-groups (e.g., family, nation), individualists show a greater degree of independence from their in-groups, which is evidenced in the importance they place on personal goals in comparison with the goals of their in-groups, the emphasis that they place on personal attitudes over in-group norms, and their social behaviour conforming relatively more to exchange theory principles of individual costs and benefits (Triandis 2001). This relatively greater emphasis on the balance of the exchange

of costs and benefits in their interpersonal relationships results in the formation and termination of their relatively more impermanent relationships being based on the shifts in these costs and benefits. According to Waterman (1984), individualists value the freedom to make choices on important issues, to take responsibility for themselves, to gain the maximum achievement with the abilities that they are endowed with, and to respect others. Schwartz (1990) states that individualism is characterised by contractual professional relationships, the importance of status achievement, and the negotiation of duties within social relationships. Individualists regard the formation of a positive self-concept as a fundamental personal characteristic that they closely associate with personal achievement and having unique rather than shared personal opinions and attitudes (Triandis 1995). For individualistic individuals, being able to openly express one's emotions and the achievement of one's goals are inherent features of personal satisfaction (Diener and Diener 1995). Furthermore, in judgements and reasoning based on the causal inferences gained from person perception, responsibility for actions is decontextualised and deemed to fall on the individual rather than the situation (Choi et al. 1999).

4.2 Collectivism

The fundamental feature of collectivism is the closer interpersonal relationships that are present within groups, which result in these groups being more cohesive. Individuals within these groups have a greater obligation to fulfil their responsibilities towards other group members (Oyserman et al. 2002). Schwartz (1990) explains that the mutual obligations and expectations that exist within the communal, collectivistic groups are determined by the statuses held by the individual members. A central feature of collectivism is the in-group vs. out-group comparison as it emphasises the outcomes, aims and values that are common to the in-group (family, clan, ethnic religions, or other groups) vis-à-vis the out-group. The social, interconnected ties within the in-group are more important than the individual, autonomous functioning of the person within that group (Triandis 1995). Rather than being a certain rigid set of values that operate within a fixed in-group, Triandis proposes that the broad range of possible in-groups dictates that collectivism, in comparison with individualism, encompasses a relatively wider set of disparate values, attitudes and behaviours that cohere according to the different social dynamics of the specific in-groups (Hui 1988). Consistent with the more social elements of collectivism, self-concept is based on group membership (Hofstede 1980), and includes characteristics such as the sacrifice of the self for others and common goals, and the maintenance of good relations (Markus and Kitayama 1991). Well-being for the collectivist is determined by successful performance in social roles and the completion of duties (Markus and Kitayama 1991). Emphasis is placed on the achievement of in-group harmony by controlling the outward expression of emotions. As meaning is contextualised, social context and the situation are deemed more significant than the individual when drawing inferences from and attributing meaning to behavioural observation (Morris and Peng 1994).

When it comes to relationality, collectivism regards membership to certain in-groups as relatively permanent and the natural way of things. Relationships within these in-groups are based on egalitarian principles, which engender a culture of generosity, but this does not extend to the out-group because of the rigid, relatively impermeable boundary that exists between them.

5 Shame and Guilt

The relative similarity between shame and guilt requires a careful assessment of their distinguishing features and how they might differ cross-culturally. As reported in Tracy et al. (2007) shame leads to more aggressive behaviour than guilt while guilt may lead to a group-based emotion of sympathy (compassion). After reviewing the evidence, Ogarkova et al. (2012) offer a comparative framework of shame and guilt based on prototypical events.

5.1 Shame

Ogarkova et al. (2012) observe that shame is elicited in response to the violation of an important social standard in which the transgressor is concerned with others' actual or imagined evaluations, which might lead to external sanctions. The feeling of being small and the desire to avoid being seen by others lead to avoidance and withdrawal behaviours. Shame is more of an intense emotion than guilt and is associated with feelings of weakness and helplessness. Finally, there is no emphasis on reparations or penance.

The relatively greater external locus of evaluations and possible sanctions characterising shame means that this emotion is likely to be more salient in the relatively more collectivistic Polish culture than an individualistic culture such as Britain. When accused of committing a social misdemeanour it could be deemed that a Polish individual might, on the basis of the emphasis on good relations and harmony, be relatively more affected by such accusations. If one is accused of wrongdoing from significant others with whom one cherishes good relations, it is easy to understand how this would exert a relatively more intense effect. Therefore, one would expect the withdrawal and avoidance tendencies associated with shame that were outlined above to be more pronounced for a relatively more collectivistic culture such as Poland. Consistent with this, Wallbott and Scherer (1995), in a large-scale cross-cultural study involving participants from 37 countries who were required to describe instances in which they had experienced emotions including shame and guilt, observed that in collectivistic cultures shame adheres more closely to the general shame profile. Wallbott and Scherer (1995) refer to this as "real" shame, and further explain that this dominates but does not exclude the presence of guilt, which in collectivistic cultures is quite distinct from shame.

5.2 Guilt

The main difference between shame and guilt is the locus of evaluations, with the former being external and the latter internal. Ogarkova et al. (2012) explain that guilt occurs when the violation of an important social norm results in remorse or regret. It is caused by internal sanctions and the individual is motivated to undertake actions in an attempt to deal with their misdemeanour, such as apologising, compensating the victim, or inflicting self punishment. There is little or no emphasis on whether the failure was public or not and there is no desire to avoid being seen.

Being relatively more individualistic, one would expect the British to have a greater degree of personal autonomy, weaker in-group relations, and place greater emphasis on personal attitudes over in-group norms. Due to a greater sense of autonomy an individualist is more likely to internalise the violation of a social norm as remorse or regret and hence experience a feeling of guilt rather than shame. Furthermore, if one views oneself as an autonomous entity with relatively weaker in-group relations and a more dismissive attitude towards in-group norms, one is more likely to internally rationalise the validity of an external accusation. Rather than accepting such accusations, as a collectivist would on the basis of a greater degree of shared identity with the accusers, an individualist is more likely to internalise the accusation and provide a judgement of their own behaviour. From the above explanations of shame and guilt it is clear that this internal rationalisation is more akin to guilt than shame. As explained by Wallbott and Scherer (1995: 481–482), individualistic "cultures may be considered as "guilt cultures", where shame turns to guilt, or where shame experiences at least involve a rather large number of guilt components". This results in similar shame and guilt experiences in individualistic cultures.

6 GRID Methodology

6.1 Background

The GRID project is coordinated by the Geneva Emotion Research Group at the University of Geneva, the Swiss Centre for Affective Sciences, the Geneva Emotion Research Group, the HUMAINE Association and Ghent University. Our present research is part of a worldwide study of emotional patterning across 28 languages, two of which are represented by two varieties (English and Chinese).

In the GRID instrument, both 24 universal prototypical emotion terms and other culturally specific emotion terms are evaluated in a Web-based questionnaire on scales representing 144 features that represent activity in six components of emotion: (a) appraisals of events, (b) psychophysiological changes, (c) motor expressions, (d) action tendencies, (e) subjective experiences, and (f) emotion regulation.

This methodology is comprehensive in its scope as it allows the multicultural comparison of emotion conceptualisations on all six of the emotion categories that have been recognised by emotion theorists (Ellsworth and Scherer 2003; Niedenthal et al. 2006; Scherer 2005).

6.2 Procedure

Participants completed the GRID in a controlled Web study (Reips 2002), in which each participant was presented with four emotion terms randomly chosen from the set of 24 and asked to rate each in terms of the 144 emotion features. They rated the likelihood that each of the 144 emotion features can be inferred when a person from their cultural group uses the emotion term to describe an emotional experience on a 9-point scale ranging from *extremely unlikely* (1) to *extremely likely* (9). Each of the 144 emotion features was presented on a separate screen, and participants rated all four emotion terms for that feature before proceeding to the next feature.

6.3 Participants

The mean ages and gender ratios of the participants for each of the emotion terms were as follows: *shame* (36 British English-speaking participants; mean age 21.3 years, 24 females); *wstyd* (22 Polish-speaking participants; mean age 23.3 years, 18 females); *guilt* (33 British English-speaking participants; mean age 22.5 years, 18 females); *wina* (17 Polish-speaking participants; mean age 21.59 years, 17 females).

7 GRID Results

7.1 Violated Laws or Socially Accepted Norms

A 2×2 Anova was performed on the means of the GRID feature "violated laws or socially accepted norms" that had two between-subjects variables: language group: Polish vs. British English; emotion: shame vs. guilt. There was a significant interaction between language group and emotion, $F\ (1, 102) = 4.76$, $p < 0.05$. Looking at Fig. 1 it can be seen that there was a non-statistically significant trend ($F\ (1, 102) = 2.32$, $p > 0.05$) showing that *guilt* was rated higher on the "violated laws or socially accepted norms" than *wina* (means of 0.83 and 0.74 for *guilt* and *wina*, respectively). The trend showing that *wstyd* was rated higher on this feature than *shame* (means of 0.84 and 0.76 for *shame* and *wstyd*, respectively) was also statistically non-significant, $F\ (1, 102) = 2.20$, $p > 0.05$.

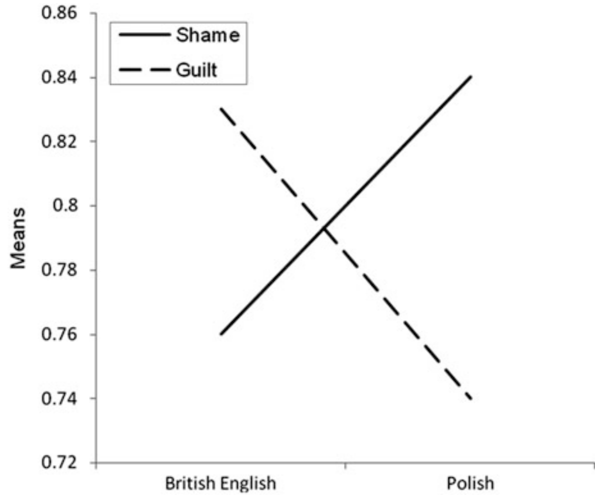


Fig. 1 Means of British English and Polish shame vs. guilt on "violated laws or socially accepted norms" GRID feature

7.2 Outward Action/Focus vs. Withdrawal/Inward Focus

The GRID features were assessed to determine those that were associated with outward action/focus vs. withdrawal/inward focus. Table (1) shows the outward action/focus features and the withdrawal/inward focus features that were selected for analysis.

(1) GRID features characterised by outward action/focus vs. withdrawal/inward focus

Outward Action/Focus	*Withdrawal/Inward Focus*
Moved toward people/things	Withdrew from people/things
Moved against people/things	Felt inhibited or blocked
Wanted to be in control of situation	Wanted to hand over initiative to somebody else
Wanted to take initiative her/himself	Wanted to submit to the situation as is
Felt an urge to be active, to do something, anything	Lacked the motivation to do anything
Wanted to move	Wanted to do nothing
Felt an urge to be attentive to what is going on	Lacked the motivation to pay attention to what was going on
Wanted to oppose	Wanted to flee
	Wanted to keep or push things away
	Wanted to prevent or stop sensory contact
	Wanted to disappear or hide from others
	Wanted to withdraw into oneself
	Wanted to break contact with others
	Wanted to run away in whatever direction

The means of the withdrawal/inward focus GRID features were subtracted from the means of the outward action/focus GRID features for each participant to create a dependent variable, the higher the value of which denoted greater outward

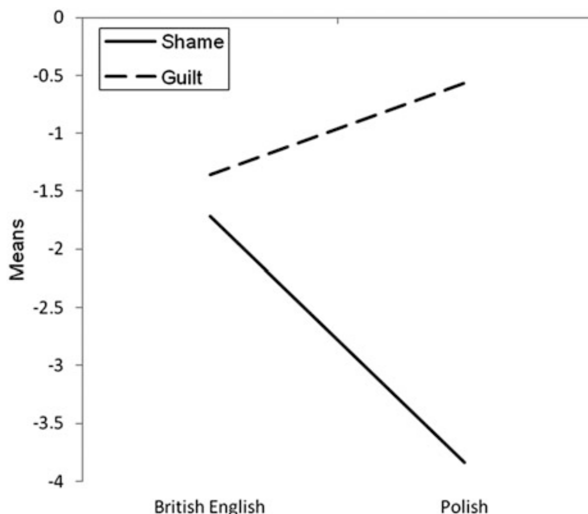


Fig. 2 Means of British English and Polish shame vs. guilt on outward action/focus vs. withdrawal/inward focus GRID features

action/focus. A 2 × 2 Anova was performed on these means that had two between-subjects variables (language group: British English vs. Polish; and emotion: shame vs. guilt). There was a significant main effect of language group, F (1, 102) = 4.95, $p < 0.05$. The British English participants rated both shame and guilt as higher in terms of outward action/focus than the Polish participants (means of −1.54 and −2.41, respectively). There was also a significant main effect of emotion, F (1, 102) = 36.64, $p < 0.001$. From Fig. 2 it can be seen that guilt is rated as having a higher likelihood of occurrence on the outward action/focus features than shame for both the British English and Polish participants (means of −1.09 and −2.55, respectively). There was also a significant interaction between language group and emotion, F (1, 102) = 51.43, $p < 0.001$. Contrasts were performed to break down this interaction. There was a significant difference in shame between British English and Polish, F (1, 102) = 14.35, $p < 0.001$. Figure 2 shows that *shame* was relatively more associated with experiences of higher outward action/focus than *wstyd* (means of −1.72 and −3.84, respectively). By contrast there was no significant difference between *guilt* and *wina* on the outward action/focus vs. withdrawal/inward focus GRID features, F (1, 102) = 0.09, $p > 0.05$. Additionally, there was a significant difference between *wstyd* and *wina* (F (1, 102) = 48.45, $p < 0.001$), but not between *shame* and *guilt* (F (1, 102) = 0.94, $p > 0.05$). Looking at Fig. 2 it can be seen that *wina* has higher ratings than *wstyd* on the outward action/focus GRID features (means of −0.57 and −3.84, respectively).

Pearson correlations were performed between the mean *shame* ratings and the mean *guilt* ratings of the British participants, and between the mean *wstyd* ratings and the mean *wina* ratings of the Polish participants for the complete profile of GRID features. The correlation between *shame* and *guilt* was 0.88 ($p < 0.001$) and the correlation between *wstyd* and *wina* was 0.76 ($p < 0.001$).

7.3 Summary of the GRID Results

There were trends showing that whereas British English *guilt* was rated relatively higher than Polish *wina* on the violation of laws or socially accepted norms, Polish *wstyd* had a higher score than British English *shame* on this dimension. Analyses performed on the means of the withdrawal/inward focus GRID features and the outward action/focus GRID features revealed significant differences in shame and guilt between the British English and Polish participants. Whereas *shame* was rated significantly higher than *wsytd* on outward action/focus, there was no difference between *guilt* and *wina* on these features. Additionally *wina* had a significantly higher rating on outward action/focus than wstyd, but there was no significant difference between *shame* and *guilt* on corresponding analyses. Finally, there was a slightly higher correlation between *shame* and *guilt* (0.88 ($p < 0.001$)) than between *wstyd* and *wina* (0.76 ($p < 0.001$)) for the complete profile of GRID features.

8 Corpus Data and Methodology

The study uses data from large corpus materials of English and Polish: the British National Corpus (BNC) and a balanced part of the National Corpus of Polish (nkjp.pl), structured along the BNC-modelled criteria and dimensions (cf. Przepiórkowski et al. 2012). Our theoretical approach at this point of the analysis combines lexical semantics and cognitive linguistics and proposes a *cognitive corpus linguistics* methodology in order to identify patterns of use and their distribution relevant to the topic.[3] We discuss insights into Polish and English cultural values, resulting from the different underlying categorizations of emotions: emotion terms in both English and Polish motivate many lexical and structural choices, and these are particularly visible in collocations. The extensive corpus data we consult are thus analyzed in terms of Cognitive Linguistics theoretical tools, such as collocation-driven categorization patterns, which appear to be most suited to their qualitative interpretation. Moreover, we generate quantitative values which reveal the frequencies of relevant emotion terms and their verbal manifestations as well as the patterns of their distributional co-occurrence with other language forms. The latter is considered to point to the salience of particular meanings most frequently co-occurring with the particular emotion terms. The physical vicinity of the lexical forms is taken to indicate their semantic overlap, as proposed by the *principle of iconicity* between form and meaning, observed in numerous cases in language structure and language use.

There are two additional sources relevant to the corpus analysis we conduct. As mentioned above, for English the British National Corpus (BNC) is used with its 100 million word collection of samples of written and spoken language from a wide

[3] See Lewandowska-Tomaszczyk and Dziwirek (2009) for an introduction of the cognitive corpus methodology and Dziwirek and Lewandowska-Tomaszczyk (2009) for an analysis of 'love' and 'hate' in English and Polish with this tool.

range of sources, designed to represent a wide cross-section of current British English. The National Corpus of Polish (NKJP) is a large collection of over one billion segments, out of which 240 million are balanced. Similar to the BNC, NKJP contains samples of written and spoken language from various styles and registers, with a structure modelled on the BNC. Due to practical reasons, selected corpus samplers were also consulted, a combination of the Microconcord and Longman data for English (18 million segments) and 10-million and 20-million segment PELCRA samplers for Polish. To obtain comparable quantitative data between the English and Polish corpora, both corpora were normalised to 100 million (NKJP) and 10 million (Samplers) segments with respect to the total data of each of the respective corpora.

The tools used in both languages are Slopeq concordancers and WordSmith concordancing tools with a set of HASK tools (Pelcra-Hask.pl), which generate sets of collocations in the form of a noun, verb, adjective or adverb for both languages.[4] An Application Programming Interface for the English version of the HASK dictionary of frequent word combinations was automatically generated from the British National Corpus (Pęzik 2013, 2014). Developed by the PELCRA group at the University of Łódź, HASK dictionaries are essentially phraseological databases meant to be used by linguists, language teachers, lexicographers, language materials developers, translators and other language professionals and casual dictionary users.

8.1 Quantitative Analysis

The English and Polish samplers contain the following numbers of segments (word list items):

LONGMAN FILES: 18 mln: 239,502 word list items
PELCRA 20 mln: 538,859 10 mln: 405,126
BNC (100 mln segments), NKJP (240 mln segments)

The frequencies of occurrence of relevant emotion terms, after the process of normalization, can be presented as follows:

(2) Frequencies of emotion terms (Microconcord and Longman corpora [M & L])

M & L (18 mln)	normalized to 10 mln		BNC 100 mln
guilt 303	ca **168**		1607
shame 325 shame* 473	ca **179** shame; ca **263** shame*		shame 1953/100 mln shame* 2709/100 mln
Pelcra 20 mln	**Pelcra 10 mln**		
(poczucie) winy	1052	439	'guilt'
wstyd (*)	1852	841,577	'shame'
wstyd 240,192,461 words/4,447/ca 185 [10 mln] ca 1850/100 mln			
wstyd* 240,192,461 words/13,852/577 [10 mln] ca 5770/100 mln			

[4] Raw frequencies and full statistics on association scores (T-test, lok-likelihood, mutual information, chi-square, Julliand's dispersion measure and range), computed for each collocational combination – not included here for reasons of space – are available from the authors on request.

The overall frequency of all emotion terms is higher in Polish than in English, which can be interpreted as typological differences between the two languages with respect to part-of-speech preference patterns rather than evidence of the linguistic preferences of the relevant speakers and writers.[5] Moreover, the Polish samplers, relative to the complete Polish corpora, contain smaller amounts of spontaneous spoken data and more numerous samples of journalistic prose and literary texts, with the latter particularly conducive to a greater prevalence of baroque, emotion and emotional language.[6] Spoken language, although clearly marking the emotional layers of meaning, will often perform this indirectly with a more constrained use of explicit emotion terms. To make this part of our methodology more consistent with the GRID method, the terms investigated in this study are primarily emotion nouns and nominal concepts collocating with verbs and with adjectives, with some observations referring to other nominal concepts used in the same context (primarily other emotion-rich expressions) to identify preferred emotion cluster patterns of the semantic prosodic type as well as other parts of speech (particularly adjectives and verbs) to cover a fuller range of related emotion concepts. The last section of the present paper contains some observations and examples drawn from the PELCRA English-Polish and Polish-English parallel corpora to provide relevant materials for the juxtaposed English and Polish discourse strategies in use.

8.2 Qualitative Analysis

8.2.1 Shame

English *shame* – BNC Verbal collocates[7]: seem, feel, bring, waste, attach, die, lose, cover, let, be should, think, could, say, know, see, can, will, get, do, would, have.

English *shame* – BNC Adjectival collocates: great, terrible, bloody, damn, real, nice, deepest, daily, awful, poor, hot, whole, full, young, able, old, social, good, other.

---

[5] See Wierzbicka (1992, 1994) and Dziwirek and Lewandowska-Tomaszczyk (2010) for a discussion of the part-of-speech based differences of expressing emotions between Polish and English. As is also found in Wierzbicka (1999) Polish emotion terms are more frequently expressed in some types of discourse as adjectives, while in their English (translational) equivalents it is the corresponding nominal structures that are preferred in the examined data as e.g., in:

Lennie dropped his head *in shame* at having forgotten./Lennie spuścił głowę *zawstydzony* tym, że się zapomniał.

He lowered his head *in shame*/Opuścił głowę *zażenowany*

And yet, in larger samples the proportions are different: BNC (100 mln segments): *ashamed* 1023, *with shame* 49, *in shame* 36, *of shame* 135; NKJP (ca 240 mln segments): Adj *zawstydzon** 787, Prep N *ze wstydem* 223, *ze wstydu* 473).

[6] Compare Bednarek (2008) for a differentiation between the language of emotions and emotional talk.

[7] The collocates for all investigated emotion terms in English and Polish are listed according to their decreasing frequencies.

(3) Sampler *shame* collocates (selection)[8]:
GUILT
MAN
ANGER
OH
FEELING
GREAT
BLUSHING
SHAMEFACED
TERROR

The concordances below present frequent clustering of the concept *shame* with *guilt*, and, less often, with *terror*, *anger* and *disgust* in the English materials as exemplified in the Sampler collocates and concordances:

(4) Concordances ***shame***

1 Commander Abigail didn't even like the area: it caused him shame and guilt to consider it, he tried not to think about it.
2 Do not think in terms of "guilt complexes" or "shame" unless you think of them as engram content
3 forgiveness or understanding or anything. But I am glad to confess that shame to your face.
4 thinking such deliberate flouting of the conspiracy of shame to be in doubtful taste.
5 There's the same flush of shame, of guilt and terror, and of cold disgust with myself.
6 I felt a deep personal guilt and shame for my country and for myself as part of that country

English shame is distributionally close to guilt with the properties of guilt prevailing (e.g., ex. 3 above, with the use of the verb *confess*, most characteristic of the *guilt* collocations).

(5) Polish ***wstyd*** **'shame' – NKJP Verbal collocates –** przyznać 'admit', być 'be', przynosić ' bring', przynieść 'bring (perfective}' powiedzieć 'say, tell', zrobić 'do', najeść 'eat up', znać 'know', czuć, 'feel', spalić 'burn up', odczuwać 'feel', przyznawać 'admit', oszczędzić 'save', poczuć 'feel', narobić 'do', palić 'burn', ogarnąć, 'overwhelm, embrace', robić 'do', pokazać 'show', ograniać 'overwhelm, embrace', przełamać 'break', kończyć 'finish, end', umrzeć 'die', wyznać 'confess', płonąć 'burn', rumienić 'blush', budzić 'wake up', mówić 'speak', wspominać 'mention, recall', pokonać 'defeat, conquer', dławić 'choke, throtle', patrzeć 'look', płakać 'cry', schować 'hide', przegrać 'lose, fail', przeżywać 'go through', wyznawać 'confess', przełamywać 'break', znieść 'stand, bear', pozostać 'keep', uciekać 'run away', wstydzić 'to be ashamed', pozbyć 'get rid of, zapomnieć' 'forget', znikać 'disappear', ukryć 'hide, conceal', doznać 'get affected', czerwienić 'redden', umierać 'die', oblać 'pour over' (metaph.), 'fail' (metaph.), skręcać 'twist, turn', piec 'bake', odczuć 'feel', wywoływać 'call forth', powodować 'cause', czerwienić 'redden', pomyśleć 'think', przemóc 'overcome', kłamać 'lie', zarumienić 'to get red', gadać 'chatter', poczerwienić 'to get red', przejmować 'take seriously', kraść 'steal', przeżyć 'live through', przezwyciężyć 'overcome', chować 'hide'

(continued)

[8] The numbers preceding the collocates are retained for those forms which fall outside of the first twenty. The structure words are omitted for the purpose of this study. The source of additional data on Nominal concepts collocating with emotion concepts are generated for English and Polish.

Polish NKJP *wstyd* 'shame' – Adjectival collocates: fałszywy 'false', żaden 'no one', wielki 'great', straszny 'terrible', czerwony 'red', głęboki 'deep', taki 'such', doroczny 'yearly', nagły 'sudden', bezsilny 'defenseless', potworny 'monstrous', ogront 'huge', niewytłumaczony 'unexplainable', pąsowy 'scarlet', zadawniony 'old', paraliżujący 'paralyzing', pomieszany 'mixed', zwyczajny 'common', gorzki 'bitter', straszliwy 'horrible', porządny 'orderly', jaki 'what a', lekki 'light', bolesny 'painful', skłonny 'prone', podwójny 'double', brudny 'dirty', cały 'whole', zbiorowy 'collectivist', moralny 'moral', gorący 'hot', wasz 'your ', seksualny 'sexual', tamten 'that', szczęśliwy 'happy', zdrowy 'healthy', silny 'strong', prawdziwy 'true', dziwny 'strange', własny 'own', ciężki 'hard heavy', powszechny 'common', jakiś 'certain', pełny 'full', osobisty 'personal', naturalny 'natural', mój 'my', ludzki 'human'

The above collocations (*shame* and *wstyd*) and those listed with *poczucie winy* 'guilt' and *guilt* are generated from the BNC and NKJP resources respectively and cover top collocates identified separately with reference to the major part-of-speech forms of the emotion terms in the consulted data, i.e., verbs and adjectives. Polish collocations in tables (6) and (9) are generated from NKJP and present top four parts-of-speech collocates: nouns, verbs, adjectives and adverbs. Collocates in tables (3, 7, 10) present relevant noun, pronoun, verb, adjective, and adverb collocates from the Samplers (Pelcra for Polish and M & L for English, respectively), while table (12) identifies most frequent Polish Pelcra Sampler patterns for *poczucie winy*.

(6) Polish NKJP *wstyd* collocations

1. rumieniec 'blush'	rumieniec___wstydu *(16)*, rumieńcem___wstydu *(7)*, rumieńce___wstydu *(2)*, wstydliwym___rumieńcem *(1)*, wstydliwa___rumieńcami *(1)*, rumieńcach___wstydu *(1)*,	28	41,633.3
4. **skrywać 'hide'**	wstydliwie___skrywane *(5)*, wstydliwie___skrywanych *(3)*, wstydliwie___skrywaną *(2)*, wstydliwie___skrywana *(2)*, wstydliwie___skrywany *(2)*, skrywany___wstyd *(2)*, skrywający___wstydliwie *(1)*, skrywając___wstyd *(1)*, wstydliwie___skrywają *(1)*, wstydliwie___skrywały *(1)*, wstydliwie___skrywano *(1)*, wstydliwą___skrywaną *(1)*, wstydliwie___skrywanej *(1)*, skrywali___wstydliwe *(1)*, wstydliwie___skrywanym *(1)*, wstydliwie___skrywał *(1)*, skrywający___wstydliwą *(1)*,	27	19,545.64
5. **przemilczeć 'leave unsaid, conceal'**	wstydliwie___przemilczane *(3)*, wstydliwie___przemilczana *(3)*, wstydliwie___przemilczany *(2)*, przemilczaną___ wstydliwie *(2)*, wstydliwie___przemilczanych *(1)*, przemilczanego___wstydu *(1)*, wstydliwie___ przemilczanej *(1)*, wstydu___przemilczał *(1)*, wstydliwie___przemilczano *(1)*,	15	9,263.41

Data in table (6) presents examples of the most common *wstyd* collocates (with their inflectional variants), in which a cluster of English emotions 'shame/timidity/ embarrassment' can be considered equivalent, i.e., *rumieniec wstydu* 'a blush of shame/timidity/embarrassment', *wstydliwy rumieniec* 'timid/embarrassed blush', *wstydliwie skrywany* 'hidden timidly/out of shame or embarrassment', *przemilczany ze wstydu/wstydliwie* 'concealed or ignored (lit. unuttered) timidly/out of shame or embarrassment'.

8.2.2 Guilt

English *guilt* – BNC Verbal collocates: feel, admit, prove, rack, establish, anger, experience, assuage, overwhelm, fear, presume, consume, cause, imply, share, suffer, play, must, may, could, come, can, take, make, would, go, will, have, do, be.

English *guilt* – BNC Adjectival collocates: sexual, collective, strong, sharp, convinced, terrible, individual, conscious, sudden, overwhelming, free, deep, german, white, certain, real, personal, human, similar, whole, full, old, national, great, good, new.

(7) Sampler *guilt* collocates (selection)
FEAR
MOTHER
PLEA
INNOCENT
MURDER
SHAME
SEXUAL
CRIME
DEFENDANT
POOR
ANGER
ANXIETY
REMORSE
SIN
CHARGE

(8)	**Concordances (*guilt**)**
5	Marie's parents had strong religious views that made her feel very guilty about having sex in their home when she and her husband had to live there
6	hurt and panic on her face. A moment ago he had experienced a twinge of guilt about the disparity in their ages
7	and still another reason to feel guilty about such wishes.
8	smoking in the workplace is becoming less and less acceptable. Many smokers feel guilty about their habit and most would like to give up.
9	Did all sons feel such guilt about their relationship with their parents.
269	Tom Ripley had one compensation, at least: it relieved his mind of guilt for the stupid, unnecessary murder of Freddie Miles.
270	He told her he forgave her, which only increased her sense of guilt, for surely there had to be something to forgive
271	But my bitterly-phrased thoughts brought no relief, only renewed tears of guilt for which I refused to seek a cause.
282	At that moment he was occupied in two ways: first on feeling guilty for telling a lie,
283	Guilt is culpable responsibility – we are guilty of some specific offence, or we may be seen by others or see them as so

Polish *poczucie winy* '(sense of) guilt' – NKJP Verbal collocates: przyznać 'admit', ponosić 'incur, suffer, be blamed for', przyznawać 'admit', obarczać 'burden, blame', udowodnić 'prove', zwalać 'blame somebody else for the Experiencer's guilt', poczuwać 'feel', zrzucić/zrzucać 'blame somebody else for the Experiencer's guilt' popełnić 'commit', obciążać 'blame somebody else for the Experiencer's guilt', zwalić 'blame somebody else for the Experiencer's guilt', przypisać 'attribute, ascribe, blame', uznać 'recognize, acknowledge', wyrządzić 'commit', odkupić 'expiate', odpuścić 'absolve', nastąpić 'follow', obciążyć 'burden, blame', przypisywać 'attribute, ascribe, blame', odkupić/odpokutować (relig.) 'to make penance for guilt', odpuszczać (relig.) 'to absolve for guilt', przerzucać 'blame somebody else for the Experiencer's guilt'

(9) Collocations *poczucie winy*

#	Kolokacja	Pasujące współwystąpienia (kliknij na frekwencję, aby wyświetlić przykłady)	Ogółem	Chi^2
1.	**umyślny 'deliberate'**	winy___umyślnej *(209)*, umyślnej___winy *(1)*, winy___umyślną *(1)*,	211	6,641,252.09
2.	**poczuć 'feel/sense'**	poczucie___winy *(1,028)*, poczucia___winy *(580)*, poczuciem___winy *(250)*, poczuciu___winy *(92)*, winy___poczucie *(7)*, winy___poczuć *(1)*, winy___poczuła *(1)*,	1959	4,123,162.09
3.	**nieumyślnej 'non-deliberate'**	winy___nieumyślnej *(32)*, nieumyślnej___winy *(1)*,	33	385,928.52
4.	**zwalać 'to blame somebody else for the Experiencer's guilt/sin'**	zwalać___winy *(23)*, zwalanie___winy *(13)*, zwalamy___winy *(2)*, zwalania___winy *(2)*, zwalał___winy *(2)*, winy___zwalać *(2)*, zwalam___winy *(1)*, zwalaniem___winy *(1)*, zwalajcie___winy *(1)*, zwalajmy___winy *(1)*, zwalaliśmy___winy *(1)*,	49	141,011.48

Materials in (9) include examples of the most common *(poczucie) winy* 'guilt (Genitive)' collocates (with their inflectional variants). The most frequent one *umyś lna wina* represents a polysemous cluster of Polish *wina* 'guilt' and *wina* 'sin' with the Eng. equivalent closer to 'sin' in the sense of 'purposeful/deliberate sin', followed closely (collocation 3) by *nieumyślna wina* 'accidental/non-deliberate sin'. The second collocate is related to the collocation *poczucie winy* 'sense of guilt', common in both languages, and the last one in (9) *zwalać winę* (colloquial) includes instances which are the most frequent in a series of similar expressions related to 'blaming somebody else for the Experiencer's guilt/sin'.

(10) Sampler *poczucie winy* collocates (selection)

WSZYSTKIEMU 'everything'
WINNIŚMY 'we owe'
BOGU 'to God'
ZBRODNI 'crime'
DUCHA 'spirit'
ŚMIERCI 'death'
PAN/A 'Lord'
ŻYDZI 'Jews'

The concordances below present a frequent co-occurrence pattern between the guilt and shame terms in Polish with a number of (Catholic) religion references and allusions.

(11) Concordances

1 te natomiast wywołują poczucie winy i lęk przed karą 'they evoke a sense of guilt and fear (lit. anxiety) of punishment'
2 Nigdy przedtem nie widziałem twarzy, z której biłoby tak wielkie poczucie winy i przerażenia. 'Never before have I seen a face with such a great sense of guilt and terror'
3 Wiem, co to jest sąd ostateczny, ponieważ mam poczucie winy. 'I am aware of what the Last Judgment is because I posses a sense of guilt'
4 Głównie ślady, echa i parę pomników; ale też głęboki smutek, gniew, negacja i poczucie winy. 'Mainly traces, echos and a few monuments, but also deep sorrow, anger, negativity and a sense of guilt'.
5 Katolicyzm od początku wpaja człowiekowi poczucie winy z powodu grzechu pierworodnego, płci, seksu. 'From5 the very beginning Catholicism entrenches a sense of guilt to human being because of the original sin, biological gender and sex'.
6 Przestępca zaczyna się bać, że zostanie zdemaskowany, wstydzi się lub ma poczucie winy. 'The offender starts to frighten that s/he will be disclosed, is ashamed or feels guilty'.
7 Wstyd i poczucie winy kazały wymazywać takiego człowieka z pamięci 'Shame and the sense of guilt made this person disappear from memory'.

Another property of the language of *guilt* in Polish is the occurrence of the negative particle *nie* 'no/not' as six most frequent segments in the sentential patterns generated for the Polish Sampler by WordSmith tools.

(12) *poczucie winy* top patterns[9]

N	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	I 'and'	I 'and'	I 'and'	I 'and'		WINNI[10] 'guilty'		BYĆ 'be'	SIĘ [reflexive]		
2	NIE [neg]	W 'in'	NIE [neg]	SIĘ [reflexive]		SIĘ 'reflexive'		WINNA 'guily'	W 'in'	NIE[neg]	
3	Z 'with'	NIE[neg]	W 'in'	ŻE 'that'	JEST 'is'	WINNY 'guilty'		I 'and'	NIE[neg]	SIĘ[ref.]	NIE[neg]

[9] The lines deleted (e.g., 9, 10) refer to homonymous forms (e.g., *winnica* 'vineyard'), which are irrelevant to the present theme.

[10] Another polysemic sense of the Adjectival form *winni* (and its inflectional variants) can refer to the sense of obligation '(they) are supposed to/obliged to'

The negative marker *nie* ‘no/not’ [neg] functions as a component reinforcing the element of ‘blaming somebody else (not Experiencer) for Experiencer’s guilt’, and is also related to some of the most frequent collocates of *wyrzuty sumienia* ‘remorse’ such as *dręczyć* ‘torment’, *powiedzieć* ‘say’, *czuć* ‘feel’, *spojrzeć* ‘look’, *zagłuszyć* ‘silence down’, *pozbyć* ‘get rid’, *zabić* (metaph.) ‘kill’.

Poczucie winy ‘guilt’ closely co-occurs with *wstyd* ‘shame’ in the Polish data, with the properties of collectivist *wstyd* prevailing as in the acts of blaming somebody else for the Experiencer’s guilt for the reasons of public stigmatization avoidance. Equally characteristic are religious contexts of *wina* ‘guilt’, particularly due to its very close polysemic links with the concept of *sin* and used invariably as a substitute of *grzech* ‘sin’ in all religious texts as e.g., in the prayer *Our Father* the English verse *Forgive us our sins* is fully equivalent to *Odpuść nam nasze winy* lit. ‘Forgive us our guilts’.

8.3 PARALLEL (Translation) CORPUS Data from English to > Polish and from Polish to > English

The last set of data concerning the emotion terms studied in the present work is generated from the PELCRA English-to-Polish and Polish-to-English translation corpora and shows a degree of approximation between the original and the target language forms.[11] It is argued in Lewandowska-Tomaszczyk (2012) that for ontological, cognitive, volitional or deontic reasons, language speakers do not use linguistic untis which would exhibit absolute identity with their conceptual intentions but rather resort to the ‘second, third, or nth best’ options in communication contexts – intercultural encounters and translation being some of them. Yet, the target form always bears some degree of resemblance to the originally intended thought and it can be proposed that the translation Target Language equivalents used in the corpora and exemplified below exhibit a stronger or weaker conceptual similarity to those used in the source language context. Furthermore, they frequently foreground the most salient properties of the source language form available in the target language repertory.

The data analysed are generated taking Polish *wstyd** and the related derived forms as search words (Verbs *wstydzić się, zawstydzać,* Nominal *zawstydzenie*, Adjectival *zwstydzony* as the search words), either in the Source (Polish) language or in the Target (Polish) language in the Polish-to-English or English-to-Polish translations.

[11] For a discussion of the processes of approximation in translation see Lewandowska-Tomaszczyk (2012).

(13) *wstyd* parallel data (English to Polish and Polish to English)
WSTYD English to Polish

Polish *wstyd* aligns with English *embarrassment* (being sorry), *bashfulness*, *timidity*, *annoyance*, *disconcertedness*.

English *shame* additionally aligns with Polish *upokorzenie* 'humiliation', *niesława* 'dishonour/ infamous', *hańba* 'disgrace', *zniewaga* 'insult'

(a)

Source text: The pitiable condition of my dupe had thrown an air of *embarrassed gloom* over all;

Target text: Opłakane położenie mej ofiary ogarnęło wszystkich *uczuciem wstydu i smutku*.

(b)

Source text: I looked at him curiously and met *his unabashed and impenetrable eyes*.

Target text: Spojrzałem z ciekawością i spotkałem jego *oczy nieprzeniknione, ale wstydu nie zdradzające*.

(c)

Source text: A look of *intense annoyance* succeeded.

Target text: Po chwili objawił się w nich wyraz *wielkiego zawstydzenia*.

(d)

Source text: Threats and even, *I am sorry to say*, blows refused to move her.

Target text: Ani groźby, ani nawet, *wstyd mi powiedzieć*, bicie, nie zdołały złamać jej oporu.

(e)

Source text: Tom was a trifle *disconcerted*.

Target text: Tomek *zawstydził się nieco*.

(f)

And Paul felt a sudden shame that he had doubted Halleck even for an instant.

I poczuł nagły wstyd, że choć przez chwilę zwątpił w Gurneya.

(g)

total victory of his astral narrative over the wretchedness and shame

Pełny tryumf jego astralnej opowieści o nędzach i hańbach świata sublunarnego.

(h)

"Now have at them, in the name of God, for a shame like this I cannot bear!"

"Teraz na nich, w imię Boże, nie mogę bowiem znieść takiej zniewagi!"

(i)

Source text: I don't think I could stand the shame.

Target text: Chyba nie zniosę takiej hańby.

(14) WSTYD > Polish to English

(a)

Source text: – Obrotny masz waść dowcip i *podobno od wstydu większy*.

Target text: You have a nimble mind, but *inclining rather to disgrace*.

(b)

Source text: *A co do wstydu*, nikogo nie zapraszam, by go pił ze mną – sam go wypiję, i da Bóg, że mi nie będzie gorzej od tego miodu smakował.

Target text: *And as to disgrace*, I ask no one to drink it with me, – I drink it alone; and God grant that it taste no worse than this mead.

(c)

Source text: W jej twarzy zabłysło szczęście, odwaga, a te porywy, walcząc ze *wstydem dziewiczym*, umalowały jej policzki w śliczne kolory różane.

(continued)

Target text: Happiness with courage gleamed in her eyes, and those impulses struggling with *her maiden timidity* painted her face with the beautiful colors of the rose.

(d)

Source text: – Z chłopką ty by inaczej gadał – *ale tej ci wstyd*.

Target text: You would talk differently with a peasant girl, *but you are afraid of her*.

(e)

Source text: Starania moje i cały ów splendor zdawały się tylko *zawstydzać*, mieszać i męczyć dziecinę;

Target text: My efforts and all that splendor seemed *to embarrass her*, confusing and tormenting the child;

(f)

Source text: W ciekawości, z jaką patrzyliśmy wzajem na siebie, tkwiła już jakaś *wstydliwość młodzieńcza i dziewicza*.

Target text: In the curiosity with which we looked at each other was hidden the undefined *bashfulness of a youth and a maiden*.

The fixed phrase *What a shame* is typically rendered in terms of pity *(Okropnie) szkoda* or bad luck *Pech*.

(15) parallel *guilt* > Polish concordances (translation directionality Polish > English and English > Polish)

(a)

Zełgaliśmy, bo *wstyd* nas żarł.

We lied – it be *guilt* devours us.

(b)

I felt a little *guilty*

Zacząłem *żałować* 'I started to regret'

(c)

Guilt-ridden

Dręczony *wyrzutami sumienia*

(d)

What lay ahead was the final death of which Lestat *was guilty*.

Oczekiwała ją wiadomość o śmierci brata, śmierci, za którą *odpowiedzialny* był Lestat.

The most characteristic property of the parallel *poczucie winy* 'guilt' concordances is a frequent translation strategy of imposing equivalence between English *guilt* and Polish *wstyd* 'shame', and furthermore between *guilt* and Polish *żałowanie czegoś* 'regret', *odpowiedzialność* 'responsibility' or *wyrzuty sumienia* 'remorse'. Polish *wstyd* is also equally frequently rendered in terms of the English prototypical lexicographic equivalent *shame*. The mutual inter-substitutability of these terms in language use, not only in translation but also in monolingual communication, presents their conceptual resemblance and provides additional evidence for the phenomenon of continuous *semantic approximation* (proposed in Lewandowska-Tomaszczyk 2012), performed by language users in interactional contexts.[12]

[12] See Lewandowska-Tomaszczyk (2012) for a more extensive discussion of *semantic approximation*, the content of approximative spaces, allowable substitutions in the spaces as well as their tolerance threshold.

9 General Conclusions

The comparison between British English vs. Polish shame and guilt on the GRID instrument features was consistent with what one would expect when comparing a relatively more individualistic vs. a relatively more collectivistic culture, respectively. The trend showing the higher ratings of *wstyd* relative to *shame* on social norm violation points to the more collectivistic Polish individuals more acutely experiencing shame because the value they place on interpersonal relationships becomes pertinent when they are accused of a social misdemeanor. More precisely, the external accusations of others, which is a key feature of shame, is particularly important to Polish individuals as the opinion of others matters more to them because of the emphasis that they place on harmonious interpersonal relationships. Further analyses on the orientation of action/focus GRID features showed that the effect of this shame, or what Wallbott and Scherer (1995) refer to as the "real" shame present in more collectivistic cultures, is a relatively greater degree of withdrawal/inward focus. In contrast, the trend showing the greater likelihood of *guilt* in response to social transgressions is consistent with the greater autonomy afforded to the more individualistic British, resulting in the internalisation of these misdemeanors as remorse or regret. In addition, the greater outward action/focus of *shame*, which is typically more characteristic of guilt, in comparison with *wstyd*, and the greater proximity of *shame* and *guilt* on the orientation of action/focus GRID features in comparison with *wstyd* and *wina* is consistent with *shame* incorporating more guilt features and with Britain, as one would expect of an individualistic culture, therefore being a relatively more "guilt culture" (Wallbott and Scherer 1995). Pearson correlations performed on the complete profile of 144 GRID features provide some, albeit not too strong, support for the relatively greater similarity between *shame* and *guilt*.

The corpus analyses show that *wstyd* is a prototypical social emotion, supposed to be hidden, but betrayed by blushing (topmost collocation above), bringing anxiety (lęk), humiliation (upokorzenie), disgrace (hańba), and is common in the shame and embarrassment (zażenowanie) cluster. *Wstyd* collocates with other emotion concepts to form clusters of the same (negative) polarity: strach 'fear', rozpacz 'dispair', smutek 'sadness', wina 'guilt', gniew 'anger', cierpienie 'suffering', and złość 'anger'.

Interestingly, further corpus analyses showed that the profile of *wstyd* overlaps with that of guilt (*poczucie winy*) in this respect, which can thus be considered a stronger emotion in Polish than in English, correlating with parts of both English *shame* and *guilt*, and additionally situated in the Catholic religious context. Consistent with the GRID data, the Polish corpus data show that in some respect Polish, unlike English, is a stronger shame than guilt culture. The Polish corpus data also revealed that Polish individuals tend to blame others for their guilt or failure. The Polish collocates (*zrzucać, mamić, zwalać, zganiać*) 'to blame sb with one's own guilt' both in more formal and in more colloquial styles are at the top of the collocation lists in Polish. In the light of the observations of Tangney et al. (2007),

who point to a correlation between shame and the other-blaming strategy when threatened with public shame or 'faceloss' (Goffman 1955), this blame-shifting is consistent with the greater salience of shame in Polish culture. The religion-based meanings in Polish referring to unjustified/sb else's/putative guilt, remorse, attributing guilt to somebody else than oneself add supporting evidence.

Another frequent, religion-conditioned, use of guilt is related to forgiving guilt, performing penance for guilt/sins (*odpokutować/odkupić*), etc., used in the lexical context of clear religious connotations. This part of linguistic meaning structures is made even more salient by the fact that the Polish form 'wina' conceptually conflates two otherwise distinct meanings in Polish, those of 'guilt' and 'sin'.[13] Interestingly enough, the Polish *wyrzuty sumienia* (remorse (of conscience)) also collocates most often with *zagłuszyć* 'to appease remorse' lit. *deafen*, and with *wina* 'guilt'. The English corpus data reveal the relationship between shame and guilt too, although guilt properties are stronger there (*guilt* is present as the first nominal collocate of *shame*) and make shame elements weaker when juxtaposed with those characteristic of guilt.

The results of the present investigation are consistent with those presented in Lewandowska-Tomaszczyk (in press), in which the online interactional patterns that were identified revealed individual and group identity traits of the online discussion commentators via distinct emotionality exponents. Lewandowska-Tomaszczyk observed that the Polish online commentators develop the projected self-construction not only on their own positive features but also on the direct reproachment directed at the interlocutor, frequently blaming the others for their guilt or lack of success. On the basis of Tangney et al. (2007), this is what one would expect of a predominantly "shame" culture. The English comments are less negatively interactant-targeted; the users employ more interactant-centered positive politeness strategies and they do not hesitate to use self-mockery, which is not as common in the Polish data. Being a "guilt" culture, this strategy is consistent with the more constructive, nonhostile approach to confrontation that Tangney et al. (2007) propose is engendered by guilt-proneness.

An attempt to account for the differences identified both in the present paper and in Lewandowska-Tomaszczyk (in press) from the perspective of the familiar distinction proposed by Hofstede (1980), which places Poles at a relatively lower, and the British – at one of the highest levels of the individualism hierarchy, brings interesting conclusions with respect to the concept of *collectivity*.[14] The Polish data emphasize either individualism or small group collectivism (e.g., a larger family circle) and distrust towards individuals from the outside. There is also more salience associated with shame in the Polish data – the scenarios in which the

[13] Other aspects of the ambiguity of the lexical form 'wina' in Polish, not relevant to this part of the discussion, although causing significant problems in the quantitative research, refer to the other sense of the form 'wina', which is equivalent to the meaning 'wine' (in Genitive Singular) in English.

[14] Recent tests we conducted with emotion display (joy, fear, anger) in gender-balanced groups do not show the essential effect of gender on the results. More research is needed on shame and guilt in this respect (Lewandowska-Tomaszczyk and Wilson in preparation).

experiencers more closely attend to what the others are thinking of them, rather than to their own guilt. Guilt is more typical of the English users who choose to attend to the relational self, derived from interactional relations with others. Additionally, guilt is directly clustered with, or frequently substitutable for remorse, shame, disgrace and humiliation with the Polish users, and responsible for the withdrawal, hiding, etc, behaviour of these users. In contrast, the energy potential of the British is more readily activated and grows possibly to compensate for their guilt. This Polish (constrained) collectivism vs. British individualism distinction offers a convincing interpretation of the findings presented by Lewandowska-Tomaszczyk (in press). Whereas the shame-proneness in the Polish online commentators leads them to blame others, the predominance of guilt in their British counterparts engenders a more constructive approach in their confrontations.

Needless to say, a larger study of emotion clusters is needed (Lewandowska-Tomaszczyk and Wilson in preparation), which would embrace other social emotions, such as pride, compassion and their associate members (vanity, sympathy, etc., respectively), to account for a more extensive portrait of collectivistic and individualistic behaviour.

References

Bednarek, M. (2008). *Emotion talk across corpora*. New York: Palgrave Macmillan.

Beer, J. S. (2007). Neural systems for self-conscious emotions and their underlying appraisals. In J. L. Tracy, R. W. Robins, & J. P. Tangney (Eds.), *The self-conscious emotions: Theory and research* (pp. 53–67). New York: Guilford Press.

Buss, A. H. (2001). *Psychological dimensions of the self*. Thousand Oaks: Sage.

Choi, I., Nisbett, R. E., & Norenzayan, A. N. (1999). Causal attribution across cultures: Variation and universality. *Psychological Bulletin, 125*, 47–63.

Diener, E., & Diener, M. (1995). Cross-cultural correlates of life satisfaction and self-esteem. *Journal of Personality and Social Psychology, 68*, 653–663.

Dziwirek, K., & Lewandowska-Tomaszczyk, B. (2009). Love and hate – Unique transitive emotions in Polish and English. In B. Lewandowska-Tomaszczy & K. Dziwirek (Eds.), *Studies in cognitive corpus linguistics*. Frankfurt am Main: Peter Lang.

Dziwirek, K., & Lewandowska-Tomaszczyk, B. (2010). *Complex emotions and grammatical mismatches*. Berlin: Mouton de Gruyter.

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, & H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 572–595). New York: Oxford University Press.

Fillmore, C. (1977). Frames-and-scenes semantics. In A. Zampolli (Ed.), *Linguistic structures processing* (pp. 55–81). Amsterdam: North Holland Publishing Company.

Fontaine, J. R. J., Scherer, K. R., & Soriano, C. (2013). *Components of emotional meaning: A sourcebook*. Oxford: Oxford University Press.

Gilbert, M. A. (1997). Prolegomenon to a pragmatics of emotion. *Proceedings of the Ontario Society for the Study of Argumentation,* Brock University, St. Catherine's. Accessed at: www.yorku.ca/gilbert/argthry/argthry/arg.../mag1997-prol-emo.pd

Goffman, E. (1955). On face-work: An analysis of ritual elements in social interaction. *Psychiatry: Journal for the Study of Interpersonal Processes, 18*, 213–231.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills: Sage.

Hui, H., C. (1988). Measurement of individualism–collectivism. *Journal of Research in Personality, 22*, 17–36.

Izard, C. E., Fantauzzo, C. A., Castle, J. M., Haynes, M. O., Rayias, M. F., & Putnam, P. H. (1995). The ontogeny and significance of infants' facial expressions in the first 9 months of life. *Developmental Psychology, 31*(6), 997–1013.

Keltner, D., & Buswell, B. N. (1997). Embarrassment: Its distinct form and appeasement functions. *Psychological Bulletin, 122*, 250–270.

Kövecses, Z. (2000). *Metaphor and emotion: Language, culture, and body in human feeling*. Cambridge: Cambridge University Press.

Levenson, R. W. (1999). The intrapersonal functions of emotion. *Cognition and Emotion, 13*(5), 481–504.

Lewandowska-Tomaszczyk, B. (2012). Approximative spaces and the tolerance threshold in communication. *International Journal of Cognitive Linguistics, 2*(2), 2–19.

Lewandowska-Tomaszczyk, B. (in press). English and Polish self-projection in the internet comments of sports events. In B. Lewandowska-Tomaszczyk, M. Kopytowska, J. Osborne, J. Schmied, & K. Yumlu (Eds.), *Languages, cultures, media*. Chambery: University of Savoy Press.

Lewandowska-Tomaszczyk, B., & Dziwirek, K. (2009). *Studies in cognitive corpus linguistics*. Frankfurt am Main: Peter Lang.

Lewandowska-Tomaszczyk, B., & Wilson, P. A. (2013). English fear and Polish strach in contrast: GRID approach and cognitive corpus linguistic methodology. In J. R. J. Fontaine, K. R. Scherer, & C. Soriano (Eds.), *Components of emotional meaning: A sourcebook* (pp. 425–436). Oxford: Oxford University Press.

Lewandowska-Tomaszczyk, B., & Wilson, P. A. (in preparation). *Emotions in Contrast*.

Lewis, M. (1995). Embarrassment: The emotion of self-exposure and evaluation. In J. P. Tangney & K. W. Fischer (Eds.), *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (pp. 198–218). New York: Guilford.

Lewis, M. (2008). Self-conscious emotions: Embarrassment, pride, shame, and guilt. In M. Lewis, J. M. Haviland-Jones, & L. Feldman Barrett (Eds.), *Handbook of emotions* (pp. 742–756). New York: Guilford.

Lewis, M., Sullivan, M. W., Stanger, C., & Weiss, M. (1989). Self development and self-conscious emotions. *Child Development, 60*, 146–156.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 20*, 568–579.

Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology, 67*(6), 949–971.

Niedenthal, P. M., Krauth-Gruber, S., & François, R. (2006). *Psychology of emotion*. New York: Psychology Press.

Ogarkova, A., Soriano, C., & Lehr, C. (2012). Naming feeling: Exploring the equivalence of emotion terms in five European languages. In P. A. Wilson (Ed.), *Dynamicity in emotion concepts. Lodz studies in language, volume 27* (pp. 247–333). Frankfurt am Main: Peter Lang.

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analysis. *Psychological Bulletin, 128*, 3–72.

Pęzik, P. (2013). Paradygmat Dystrybucyjny w Badaniach Frazeologicznych. Powtarzalność, Reprodukcja i Idiomatyzacja. *Metodologie Językoznawstwa. Ewolucja Języka, Ewolucja Teorii Językoznawczych*. Piotr Stalmaszczyk (red.). Wydawnictwo Uniwersytetu Łódzkiego.

Pęzik, P. (2014). Graph-based analysis of collocational profiles. In V. Jesensek (Ed.), *EUROPHRAS - 2012 Maribor. Phraseology and culture. Abstracts*. Maribor: Filozofska fakulteta.

Przepiórkowski, A., Bańko, M., Górski, R., & Lewandowska-Tomaszczyk, B. (Eds.). (2012). *Narodowy Korpus Języka Polskiego*. Warszawa: PWN.

Reips, U.-D. (2002). Standards for internet-based experimenting. *Experimental Psychology, 49*, 243–256.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information, 44*, 693–727.
Schwartz, S. H. (1990). Individualism–collectivism: Critique and proposed refinements. *Journal of Cross-Cultural Psychology, 21*, 139–157.
Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology, 52*, 1061–1086.
Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Oxford: Blackwell.
Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). What's moral about the self-conscious emotions? In J. L. Tracy, R. W. Robins, & J. P. Tangney (Eds.), *The self-conscious emotions: Theory and research* (pp. 21–37). New York: Guilford Press.
Tracy, J. L., & Robins, R., W. (2004). Putting the self into self-conscious emotions: A theoretical model. *Psychological Inquiry, 15*, 103–125.
Tracy, J. L., Robins, R. W., & Tangney, J. P. (Eds.). (2007). *The self-conscious emotions: Theory and research*. New York: The Guilford Press.
Triandis, H. C. (1995). *Individualism and collectivism*. Boulder: Westview Press.
Triandis, H. C. (2001). Individualism-collectivism and personality. *Journal of Personality, 69*, 907–924.
Wallbott, H. G., & Scherer, K. R. (1995). Cultural determinants in experiencing shame and guilt. In J. P. Tangney & K. W. Fischer (Eds.), *Self-conscious emotions* (pp. 465–487). New York: Guilford Press.
Waterman, A. S. (1984). *The psychology of individualism*. New York: Praeger.
Wierzbicka, A. (1992). *Semantics, culture, and cognition: Universal human concepts in culture-specific configurations*. New York/Oxford: Oxford University Press.
Wierzbicka, A. (1994). Emotion, language, and "cultural scripts". In S. Kitayama & H. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 130–198). Washington, DC: American Psychological Association.
Wierzbicka, A. (1999). *Emotions across languages and cultures: Diversity and universals*. Cambridge: Cambridge University Press.

Corpora

English Microconcord and Longman Samplers
Polish PELCRA Samplers
The British National Corpus (BNC): http://www.natcorp.ox.ac.uk/ (100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of current British English, both spoken and written).
The National Corpus of Polish nkjp.pl

Translating Freedom Between Cultures and Ideologies: A Comparative Analysis of the Translation of Keywords in Galatians

Sarah Buchanan

Abstract This paper presents a comparative analysis of Bible translations in German, French, Spanish and English targeted towards a range of cross-confessional audiences. It focuses on the key word translation of concepts such as freedom and slavery. It examines the translator's choices and the pragmatic implications of these decisions for readers of translations of the Bible. The case study centres on the concepts of freedom and slavery in Paul's letter to the Galatians with an intercultural corpus of 16 translations. Authorised and widely accepted translations in English, German, French and Spanish such as the *Lutherbibel* (1984) and *Reina Valera* (1989) are compared with new competing translations, such as the *New Living Translation* (Tyndale House Publishers, Carol Stream, 2007), *The Message* (Peterson, The Message: the new testament in contemporary English. NavPress, Colorado Springs, 2005), and *Die Volxbibel* (2005). This study draws from the fields of Pragmatics, Translation Studies and Theology, to provide a unique cross-cultural examination of Galatians, and of sacred translation. It is found that the choices of the translator of sacred texts are not merely linguistic choices, but rather they are often rooted in various ideological and theological positions.

Keywords Bible translation • Key word translation • Pragmatics • Galatians • Cross-confessional • Freedom • Theological concepts • Sacred translation

Abbreviations

BFC La Bible en français courant (1996)
BNT BibleWorks NT (Novum Testamentum Graece, Nestle-Aland 27th Edition)

S. Buchanan (✉)
Translation and Interpreting School of Modern Languages, Queen's University Belfast, Belfast, UK
e-mail: sbuchanan03@qub.ac.uk

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2,
DOI 10.1007/978-3-319-06007-1_8, © Springer International Publishing Switzerland 2014

CJB	Complete Jewish Bible (1998)
EIN	Einheitsübersetzung der Heiligen Schrift (1980)
ESV	English Standard Version (2007)
LIT	La bible de la liturgie (1980)
LSG	La Bible: Louis Segond (1910)
LUT	Lutherbibel (1984)
MSG	The Message (2005)
MT	My Translation
NIV	The New International Version (2011)
NLT	New Living Translation (2007)
NT	New Testament
PDT	La Biblia: La Palabra de Dios para Todos (2005)
RVA	La Santa Biblia Reina-Valera Actualizada (1989)
SGB	Sagrada Biblia (2011)
VLX	Die Volxbibel (2005)
VCE	The Voice New Testament (2011)

1 Introduction and Methodology

> Solidity, indeed, becomes the pen
> Of him that writeth things divine to men:
> But must I needs want solidness, because
> By metaphors I speak? Were not God's laws,
> His gospel laws, in olden time held forth
> By types, shadows and metaphors? [. . .]
> (The Pilgrim's Progress, John Bunyan)

John Bunyan writes here in defence of his allegorical style and reminds us of the heavy weight of responsibility that presses down upon those dealing with a text considered sacred. He experienced opposition from those who expected straight forward didactical writing on the Scriptures and disliked his metaphorical style. Bunyan maps out ancient concepts and Christian experience in terms of physical places and metaphorical narratives. He embodies concepts such as vanity, death, and heaven in places such as *Vanity Fair*, *the River of Death* and *Celestial City*. Readers of Pilgrim's Progress can turn the pages and look at the map to follow Christian's journey and perhaps relate it to their own beliefs and experiences in a more tangible way. Likewise, translators of the Bible today face the task of relating sacred concepts to their audiences' convictions and experiences. In this paper we will observe the re-mapping of concepts central to Christian belief in the twenty-first century, focusing on the ideas of freedom and slavery. We will explore old maps[1] of ancient concepts: authorised translations of the Bible in German, Spanish, French and English which have been prevalent for at least one century, and new

[1] For the purpose of this paper "mapping" is used to describe the representation of a keyword from an ancient text (the Bible) in a new situational context, by means of interlingual translation, taking into account cultural and temporal factors.

maps: modern translations, written since 1980, for different audiences in German, Spanish, French and English. By analysing translations for various audiences, the aim is to explore the journey of keywords between different audiences and reveal the pragmatic implications of translation choices.

In exploring the journey of keywords, it is important to recognise the lack of commensurability between the conceptual worlds of different cultures and language systems. In linguistic theory over the second half of the twentieth century there has been rivalry between those propagating either linguistic universality or linguistic relativity. Sapir and Whorf, on one end of the scale, proclaimed that all language was culture-bound, whereas Chomsky, in contrast, asserted that human beings had a "rich and invariant conceptual system prior to any experience" (1987: 22). Anna Wierzbicka, in her work *Cross-Cultural Pragmatics* argues that: "cross-linguistic and cross-cultural variations are not minor but colossal" and she assumes there to be "in the neighbourhood of (only) 60 very simple universal primitives such as I, you, someone, something, want, don't want (. . .)" (Wierzbicka 2003: 3). In this assertion she therefore complicates the view of translation as a simple act of transfer between cultural systems.

Little work has been carried out in the context of corpus linguistics and pragmatics wherein biblical concepts have been directly compared between languages. Piotr Blumczyński (2006) carried out a corpus based study of markers of ideology in English Bible translations in his work *Doctrine in Translation: The Doctrine of the Trinity and Modern English Versions of the New Testament* and David Bell (2010), in his work *A Comparative Analysis of Formal Shifts in English Bible Translation* undertook a comparative analysis between biblical texts (in Hebrew and Greek) and their rendering in ten English Bible translations. The aim of this study is to open up to a cross-cultural understanding of Galatians, both of the concepts within the case study and of the journey of concepts between various contexts of interpretation, as various translators make different choices. To return to Wierzbicka, she writes: "For the modern Anglo reader of the Bible, a cross-cultural commentary is not an optional extra, but a necessity [. . .] it can be an effective tool of interpretation" (2004: 575).

1.1 Keywords as a Theoretical Framework

Why use keywords as a tool to examine translations? Keywords are the skeletons of texts and therefore useful in any textual or conceptual analysis. Piotr Blumczyński describes keywords as "lexical items whose sense plays a particularly strong dialectical relationship with the sense of totality of the text in which they occur" (2007: 90). Keywords are not arbitrary, but those words that play an important role in the overall ideas and arguments within a text, words that are noticed and become central to different types of discourse, in this instance contemporary religious discourse. Raymond Williams (1983) in his work *Keywords: A Vocabulary of Culture and Society* writes about keywords in two senses: "These are significant,

binding words in certain activities and their interpretation; they are significant, indicative words in certain forms of thought" (1983: 15). In Paul's letter to the Galatians there are a number of keywords which have become common currency in both subsequent Christian thought and secular philosophy, words which have taken on various connotations by different groups. Williams outlined his reasoning for an investigation into words and their relationships to other words "because the problems of meanings seemed [. . .] inextricably bound up with the problems it was being used to discuss" (1983: 15). Just as the various senses of vocabulary can reveal attitudes and beliefs, it is valuable to uncover the relationships between words in a larger framework.[2] Williams describes his experiences of coming back to England after the Second World War and noticing the development of connections between words, different from that which he remembered 4 years previous when leaving England:

> I began to see this experience as a problem of vocabulary [. . .] the explicit but as often implicit connections (between terms) that people were making, in what seemed to me, again and again, particular formations of meaning – ways not only of discussing but at another level of seeing many of our central experiences (Williams 1983: 15).

In the light of Williams' comments on connections between and around keywords, the analysis carried out in this paper involves not only a study of the single lexical item chosen to render a keyword, but rather an awareness of the translation aims and an examination of the immediate context around the lexical item, such as prepositions, grammatical features and tone of the sentence, and how one concept relates to another, for example how the translation of *ἐλευθερία* (most often rendered freedom) and *δουλεία* (most often rendered slavery) relate to one another.

For the purpose of this paper context is understood in different dimensions, that come together to make a text meaningful. Kecskes (2010: 6) writes that "what happens in communication is that context encoded in the utterances 'matches' with the actual situation context, and their interplay results in what we call 'meaning'" (Kecskes 2010: 6). In biblical texts, the "context encoded in the utterances" consists of the words and phrases around the rendering of keywords, "the actual situational context" relates to the context of interpretation, ie. the aims of the translation and identity of the audience, if known. A third context, information which comes to light from biblical studies, may be taken into consideration, as this knowledge often accompanies the reader's world as they approach the text. In Table 1, the three contexts are displayed alongside an example of questions which arise in analysis of keywords relating to them. These questions have been applied to the verse Galatians 3:28 in the Complete Jewish Bible (1998), to provide an example of the methodology of analysis:

> There is neither Jew nor Gentile, neither slave nor freeman, neither male nor female; for in union with the Messiah Yeshua, you are all one (Galatians 3:28, CJB 1998).

[2] Graham Allen (2000) reinforces this in his work Intertextuality where he writes, "the meanings we produce and find within language, then, are relational; they depend upon processes of combination and association within the differential system of language itself. This relational aspect of language cannot be avoided or overcome" (Allen 2000: 10).

Table 1 Contexts

Three contexts	Questions
Theological context	How does the socio-historical context of the book relate to the rendering of keywords?
	Is there a reflection of ideological/theological interpretations, either or present, reflected in the translation choices? (i.e. Luther's notion of freedom, or Liberation Theology)
Immediate context (Context encoded in the utterances)	How are relationships between keywords rendered (i.e. "slave" and "freeman")?
	What grammatical and syntactical features surround the keyword (i.e. neither, nor)?
	How does this occurrence relate to other occurrences in the text?
Context of interpretation ("Actual situational context")	What are the known translation aims? (to relate to interpretive portrait of Jewish belief by Christians or Jew/Christian relations)
	Who is the intended audience? (Messianic Jewish audience at the end of the twentieth century)
	What are the pragmatic implications of rendering keywords in a particular way? (i.e. choosing "freeman" over "free")
	How do these pragmatic implications compare with that of other translations?

In these three contexts (theological context, immediate context and context of interpretation), three disciplines come together in the examination of the journey of keywords in Bible translation. In sacred translation, the theological context cannot be ignored and draws insights about the text and its reception from Biblical Studies. The other two contexts (the immediate context and the context of interpretation) draw insights from both linguistics and translation studies. Therefore three disciplines come together as the rendering of keywords is examined.

1.2 Case Study: Galatians

In order to explore the journey of concepts between different Bible translations, I have selected one book as a case study for analysis: the book of Galatians in the New Testament. Galatians[3] was written to an emerging Christian community in the Graeco-Roman Empire between CE 50 and CE 60 and is attributed to the apostle Paul. Although there is some debate over the extent of this region, it has generally been agreed by most commentators that Galatians was written to "Christians of Gaulish or Celtic descent whose churches were located in Asia Minor" (Longnecker

[3] The word from which Galatians is derived, *Γαλάται*, was used interchangeably with Κελται in classical Greek, relating to the Latin *Celtae* (Celts), or *Galli* (Gauls).

1990: lxiii). The early Christian community in Galatia was ethnically diverse, composed of believers from two Jewish and Gentile backgrounds. Paul wrote the letter into a situation of conflict. His letter centred around the relationship between two groups and their understanding of the Christian message.

In the opening he writes, "I am astonished that you are so promptly turning away from the one who called you in the grace of Christ and are going over to a different gospel" (Galatians 1:6, NJB 1985). At the outset of writing this letter there was a *gospel* but then *another gospel*, two different understandings of a key concept, out of which conflict arose. Relationships between words are embedded within real life relationship and ways of seeing the world. Paul's opponents, according to the letter, tried to persuade the new Gentile believers that they needed to obey Jewish customs in order to be accepted by God and to enjoy the promises of God from the Hebrew Scriptures. Paul countered these arguments, and explained the crux of the gospel message. Therefore this book contains many key concepts to Christianity and provides a platform from which to explore the journey of concepts from the cultural and historical moment in Galatia in the first century CE to the various target audiences in the late twentieth and early twenty-first century.

The expectations and comprehension of modern-day readers no doubt takes a great departure from that of the original audience. The utterances within this first century book created "expectations which guid(ed) the hearer towards the speaker's meaning" (Wilson and Sperber 2004: 607) as the letter was read out loud in a group setting, in a context of conflict, where the letter was addressed specifically to the Galatian community. In the twenty-first century readers come, often as individuals, with varying expectations to the text and construct certain ideas about both explicit and implicit information within the text, while working out the intended implications. These implications are very important for those readers of the Bible who consider it as the sacred Word of God, and therefore, the implications of the translations of the book have been heavily debated for centuries. Here we can draw from Sperber and Wilson's relevance theory, wherein the "identification of explicit content [is considered] as equally inferential, and equally guided by the Communicative Principle of Relevance, as the recovery of implicatures" (ibid: 607) and whereby "Comprehension is an on-line process, and hypotheses about explicatures, implicated premises and implicated conclusions are developed in parallel against a background of expectations (or anticipatory hypotheses) which may be revised or elaborated as the utterance unfolds" (ibid: 608). The expectations of the readers, in the case of Bible translation, often depends on the confessional group to which one belongs, and the expectations may be either appropriated or refuted according to the strategy of the translator.

In order to explore the pragmatic implications of utterances in translation, this paper focuses on one dichotomy within Galatians: *ἐλευθερία* (freedom) and *δουλεία* (slavery). This dichotomy is part of a larger research project examining eight keywords in Galatians. There are many binary oppositions within the Pauline corpus and this traditional antithetical way of framing the world was not unique

Table 2 Keyword dichotomies in Galatians with most common rendering in English

πνεῦμα (spirit)	σάρξ (flesh)
χάρις (grace)	νόμος (law)
εὐλογία (blessing)	ἀνάθεμα (curse)
ἐλευθερία (freedom)	δουλεία (slavery)

to Paul, but rather was employed by many classical writers, such as Aristotle and Plato.[4]

The antithetical pairings laid out in Table 2 have been chosen because of their frequency[5] within Galatians and their centrality to Paul's argument, yet admittedly the act of selection is to some extent interpretive. The keywords are metaphors that Paul drew from both Graeco-Roman and Jewish traditions. Paul explained the divine in terms of structures and ideas already present within Graeco-Roman society and Jewish tradition. In this paper, the focus will be upon one opposition and how it is mapped in Bible translations: *ἐλευθερία* (freedom) and *δουλεία* (slavery). Theologian Luke Timothy Johnson (2012 in Bird 2012: 84) places this binary opposition within what he describes as economic discourse, relating to first century experience:

> In economic language, the condition of distance from God is expressed in terms of slavery [. . .]; God's action is expressed as redemption [. . .]; Christ's death as a ransom (1 Tim. 2:6) that "purchases" believers [. . .]; the result is freedom.

Ελευθερία, often rendered in English as 'freedom' or 'liberation', stands in opposition to *δουλεία* (slavery), and frequently acts as a synonym for other keywords within Galatians and the New Testament. In order to adequately analyse translations of *ἐλευθερία* and *δουλεία* in Galatians, a thematic analysis of the opposition will be carried out, to look at patterns of where it occurs within the original Greek text., before comparing translations within the corpus.

[4] Historian Brigitte Kahl describes the method of categorization:

> They were typically organized in two oppositional columns, of unequal weight and value. The items listed on one side were complementary to each other in some way and, at the same time opposed and superior to their counterparts in the other column. (Kahl 2010: 17)

[5] Frequency of keywords in Galatians

Keywords	Number of Occurrences in Galatians
πνεῦμα (MT: spirit)	19
σάρξ (MT: flesh)	18
χάρις (MT: grace)	9
νόμος (law)	32
εὐλογία (blessing)	2
ἀνάθεμα (curse)	8
ἐλευθερία (freedom)	10
δουλεία (slavery)	9

1.3 Intercultural Corpus

In order to examine the journey of the dichotomy *ἐλευθερία* (freedom) and *δουλεία* (slavery) we are going to think of translation in terms of re-mapping, and examine the shift between what will be referred to as 'old maps' and 'new maps'.

Bible translation has been dominated largely by 'authorised' versions in English, Spanish, German and French over the past couple of 100 years. These translations have been passed down through tradition and gained prominence above other translations in use and visibility. There are four 'old maps' to be considered in the analysis of *ἐλευθερία* and *δουλεία*. The *New Revised Standard Version* (1989) draws from the King James' tradition of the English Bible, with the first King James Version published in 1611. The *Reina-Valera Actualizada* (1989) is an updated version of the *Reina Valera,* the first printed Bible in Spanish, translated by the Lutheran Casiodoro de Reina and published in 1569. The *Lutherbibel* (1984) is a revised version of Martin Luther's translation, the first complete printed Bible in the German language in 1534. The *Louis Segond* was translated by Swiss theologian Segond in 1880, and published in 1910. It has not since been updated.

The 'new maps' to be considered vary in purpose, audience and translation strategy. They have not gained the same authority or visibility as the 'authorised' versions. *La Bible de la liturgie* (1980) and *Sagrada Bíblia* (2011) serve liturgical purposes, whereas *Die Volxbibel* (2005) and *The Message* (2005) are communicative translations which aim to make the Bible more palatable for a non-religious audience. *La Bible en français courant* (1996) and *La Palabra de Dios para Todos* (2005) are both missional Bible translations published by Bible societies. The *New Living Translation* (2007) is an evangelical translation, whereas *the Complete Jewish Bible* (1998) is aimed specifically toward Messianic Jewish communities, and the German *Einheitsübersetzung* (1980) tends to be used by Roman Catholic churches in Germany.

1.4 A Note on My Translation

As this is an intercultural corpus, I have provided translations alongside words in Greek, German, Spanish and French. These translations have been written for the purpose of clarity for readers. My strategy was to keep close semantic correspondence rather than write smooth stylistic translations. Where the translation is my own rather than a published translation, I have inserted the abbreviated letters MT (My Translation). This paper discusses the journey of concepts between audiences of various ideological and cultural backgrounds, therefore, it is imperative that readers do not see my translations as a transparent window into the translated texts of the corpus, but rather as an aid, after first examining the foreign language text and discussion around it. As Wierzbicka writes: "cross-linguistic and cross-cultural variations are not minor but colossal" (1987: 22), therefore, it is important not to assume commensurability when viewing a text and its translation.

2 Freedom and Slavery: Ancient and Modern Sites of Interpretation

Having looked at the actual situational context, the corpus, we now turn to the immediate context of the Greek text and examine patterns of where the binary opposition occurs, before exploring occurrences in other corpora and finally comparing translations.

2.1 Thematic Analysis of ἐλευθερία and δουλεία in Galatians

In the context of Galatians, I have divided Paul's uses of *ἐλευθερία* (freedom) into three categories: ontological, ethical and legal. I have laid out the occurrences of each of these three groupings in Tables 3 and 4.

2.1.1 ἐλευθερία as Identity

Firstly, there is a sense that *ἐλευθερία* (freedom) represents an identity and a state of being: this is the ontological sense of the word. In Galatians 2:4 Paul equates his opponents to enslavers, who spy on the freedom/liberation that the believers have in Christ Jesus. He writes: "παρεισηλθον κατασκοπησαι τὴν ελευθερὶαν ἣν ἑχομεν ὲν Χριστω'Ιησου, ὶνα ἡμας καταδουλώσουσιν"[6] (Galatians 2:4, BNT). Ελευθερία is

Table 3 Corpus of translations

English	German	Spanish	French
NLT (2007)	LUT (1984)	RVA (1989)	LIT (1980)
MSG (2005)	EIN (1980)	PDT (2005)	LSG (1910)
CJB (1998)	VLX (2005)	SGB (2011)	BFC (1996)
NRS (1989)			

Table 4 Categories of ἐλευθερία in Galatians

Category of ἐλευθερία	No. of occurrences	References
Ontological/pertaining to identity	4	Galatians 2:4
		Galatians 4:23
		Galatians 4:26
		Galatians 4:30
Ontological & ethical	1	Galatians 5:1
Ethical	2	Galatians 5:13
Legal/social status	3	Galatians 3:28
		Galatians 4:22
		Galatians 4:23

[6] This is rendered in the NIV (2011) as: "some false believers have infiltrated our ranks to spy on the freedom we have in Christ Jesus and to make us slaves" (Galatians 2:4).

something of divine origin and a mark of a true believer. Later, in Galatians 5:1, Paul writes "Τῇ ἐλευθερίᾳ ἡμᾶς Χριστὸς ἠλευθέρωσεν · στήκετε οὖν καὶ μὴ πάλιν ζυγῷ δουλείας ἐνέχεσθε" (Galatians 5:1, BNT).[7] A status of freedom has been achieved through Christ. In both these examples *ἐλευθερία* (freedom) is contrasted with *δουλεία* (slavery), where *ἐλευθερία* has a divine origin and δουλεία is something sought after or brought about by humans.

Paul grounds the believers' identity in the concept of *ἐλευθερία* (freedom) by looking at their recent experiences in Galatians 2:4 and Galatians 5:1. He develops this argument further by drawing upon Hebrew Scriptures; therefore, the Galatian believers are joined with Jewish believers in this state of *ἐλευθερία*. In chapter four of Galatians, he relates the identities of *ἐλευθερία* and *δουλεία* to the promises given to Abraham in the Hebrew Scriptures. Abraham, one of the main fathers of the faith in Judaism, received the following promise from God, written in Genesis 12:2–3a (NIV 2011):

> I will make you into a great nation,
> and I will bless you;
> I will make your name great,
> and you will be a blessing.
> I will bless those who bless you,
> and whoever curses you I will curse.

It is written that Abraham was promised many descendants yet he did not have children at the time of this promise, so he took for himself a slave woman because he believed his wife could not conceive in her old age. However, after having a son – Ishmael – to the slave woman, he had another son, Isaac, to his wife Sarah, and Isaac has since been considered the true heir to the promises of God, by both Jews and Christians. According to the arguments of Paul's opponents, however, the new Gentile believers needed to obey Jewish customs in order to become full heirs of the Abrahamic promises and therefore fully belong to the group of believers. Paul counters this argument and re-interprets the story of Abraham in the light of the situation in Galatia. He writes: "γέγραπται γὰρ ὅτι Ἀβραὰμ δύο υἱοὺς ἔσχεν, ἕνα ἐκ τῆς παιδίσκης καὶ ἕνα ἐκ τῆς ἐλευθέρας." (Galatians 4:22, BNT), which the NRS (1989) renders as: "For it is written that Abraham had two sons, one by a slave woman and the other by a free woman." He outlines the social status of Abraham's wives, as enslaved or free, and relates this to the current identity of the believers.

In Galatians 4:26 Paul aligns the two mothers, Hagar and Sarah with two Jerusalems – an earthly and a heavenly Jerusalem, he writes that Sarah corresponds to the free Jerusalem, as do the believers at Galatia. He traces the identity of the *ἐλεύθερος* (the liberated/the free) back to Abraham, therefore drawing Gentiles into the promises made to the Jewish peoples, and making connection between the Jewish Scriptures and the early Christian communities. *Ελευθερία* (freedom) and its' antonym *δουλεία* (slavery) are widely used to tell Jewish history, and present an

[7] The NJB (1985) translates this verse "Christ set us free, so that we should remain free. Stand firm, then, and do not let yourselves be fastened again to the yoke of slavery" (Galatians 5:1, NJB 1985).

important leitmotif in this re-telling throughout Scripture and church history. In the Old Testament, the Exodus from slavery in Egypt constitutes one of the most significant narratives and remains central to both Christian and Jewish belief today. Before the Ten Commandments were given to Moses, these words are recorded: "I am the Lord your God, who brought you out of the land of Egypt, out of the house of slavery" (Exodus 20:1, NIV 2011). Christian theologians have termed the coming of Christ as the "Second Exodus", and both *ἐλευθερία* (freedom) and *δουλεία* (slavery) occur frequently as themes in the New Testament, particularly in Paul's letters.

2.1.2 ἐλευθερία as an Ethical Entity

Secondly, *ἐλευθερία* (freedom) has an ethical sense in Paul's writing. It relates to behaviour of the believers and how they live in community. Although they have received a status and inheritance as free children, Paul warns them and exhorts them in their lifestyle. He explains why Christ has redeemed them in Galatians 5:1: for *ἐλευθερία*, for this liberated status, and warns them not to return to a yoke of slavery. Later, in Galatians 5:13, toward the close of his letter, Paul writes: "Ὑμεῖς γὰρ ἐπ᾽ ἐλευθερίᾳ ἐκλήθητε, ἀδελφοί · μόνον μὴ τὴν ἐλευθερίαν εἰς ἀφορμὴν τῇ σαρκί, ἀλλὰ διὰ τῆς ἀγάπης δουλεύετε ἀλλήλοις" (Galatians 5:13, BNT).[8] In every other occurrence in Galatians apart from this one, *ἐλευθερία* is the positive quality or status coming from the divine, contrasted with the negative status of those outside the group of true believers, either enslavers or enslaved. Yet in Galatians 5:13 Paul turns these values on their head and asks the believers not to use their *ἐλευθερία* as an occasion for the flesh, but rather to be enslaved to oneanother in love. *ἐλευθερία* and *δουλεία* are used in an ethical sense, in an exhortation on how to live. Here *δουλεία* is used in a positive sense, to encourage selflessness in relationship with others.[9] *ἐλευθερία* therefore, in Paul's writing, represents not only a lack of constraints, as may be imagined by today's reader, but impacts upon lifestyle.

---

[8] The ESV (2007) renders Galatians 5:13: "For you were called to freedom, brothers. Only do not use your freedom as an opportunity for the flesh, but through love serve one another."

[9] Theologian Douglas A. Campbell (2012) points to the historical context and the contemporary thinkers of Paul in an effort to interpret the term ἐλευθερία in this instance. He bemoans the gap between our post-enlightenment minds and the era in which Paul wrote his letters.

> Paul [. . .] is not operating with modern notions of causality and freedom. [. . .] He is almost certainly informed distantly by Greek-speaking philosophers [. . .] In an embodied, complex and relational situation, freedom is not a matter of sheer choice [. . .] but of an incremental creation of new possibilities for bodily action that must be learned and internalized. (Campbell in Nanos 2012: 132)

Table 5 ἐλευθερία within other corpora

Corpus	No. of occurrences of ἐλευθερία
Galatians	10
Entire Pauline corpus (including Galatians)	33
Gospels	18

2.1.3 ἐλευθερία as a Legal Status

Finally, Paul uses *δουλεία* and *ἐλευθερία* to relate to the legal status and societal position of believers. He employs these terms to describe the status of Abraham's wives in the Jewish Scriptures, and then he relates δουλεία and ἐλευθερία to the present circumstances of the community in Galatia wherein old boundaries have been abolished in Christ. He exhorts them to work for a united identity and maintain unity among believers in Galatia, despite social, economic, religious or gender differences. He writes: "οὐκ ἔνι Ἰουδαῖος οὐδὲ Ἕλλην, οὐκ ἔνι δοῦλος οὐδὲ ἐλεύθερος, οὐκ ἔνι ἄρσεν καὶ θῆλυ• πάντες γὰρ ὑμεῖς εἷς ἐστε ἐν Χριστῷ Ἰησοῦ" (Galatians 3:28). The English Standard Version (2007) renders this: "There is neither Jew nor Greek, there is neither slave nor free, there is no male and female, for you are all one in Christ Jesus" (Galatians 3:28, ESV 2001).

2.2 Occurrences of ἐλευθερία Within Other Corpora

There are ten occurrences of *ἐλευθερία* (freedom) within my case study, and I have studied their translation within a corpus of fifteen translations. However, *ἐλευθερία* (freedom) and *δουλεία* (slavery) constitute an important motif in other corpora, for example within both the Pauline corpus[10] and other New Testament books, such as the Gospels: Matthew, Mark, Luke and John (Table 5).

Theologian Rudolph H. Blank confirms that "the Greek word ἐλευθερία [...] occurs with greater frequency in Paul's Letter to the Galatians than in any other New Testament writing" (1994: 268). Within the Pauline corpus, there are appearances of *ἐλευθερία* and *δουλεία* in Romans, 1 Corinthians, Ephesians and Philippians. In Romans 6:18 Paul describes the believers' position before God, in terms of being set free from sin: "ἐλευθερωθέντες δὲ ἀπὸ τῆς ἁμαρτίας"[11] (BNT). Echoes of

[10] Thirteen epistles in the New Testament are generally attributed, by scholars, to the apostle Paul. Some of these letters were addressed to Christian communities known to Paul, and have the two-fold aim of dealing with specific crises they faced and outlining core beliefs. These encompassed Romans, 1 Corinthians, 2 Corinthians, Galatians, Ephesians, Philippians, Colossians, 1 Thessalonians, and 2 Thessalonians. Moreover, he wrote four letters to individual leaders in the early Christian Church which are included in the New Testament Canon. These letters (1 Timothy, 2 Timothy, Titus and Philemon) were addressed to the individuals Timothy, Titus and Philemon respectively.

[11] MT: after you had been set free from sin (Romans 6:18).

Table 6 Use of ἐλευθερία and δουλεία in the Gospels

John 8:33–36 (BNT)	ἀπεκρίθησαν πρὸς αὐτόν · σπέρμα Ἀβραάμ ἐσμεν καὶ οὐδενὶ δεδουλεύκαμεν πώποτε · πῶς σὺ λέγεις ὅτι **ἐλεύθεροι γενήσεσθε**; ἀπεκρίθη αὐτοῖς ὁ Ἰησοῦς · ἀμὴν ἀμὴν λέγω ὑμῖν ὅτι πᾶς ὁ ποιῶν τὴν ἁμαρτίαν **δοῦλός ἐστιν τῆς ἁμαρτίας**. [...]**ἐὰν οὖν ὁ υἱὸς ὑμᾶς ἐλευθερώσῃ**, ὄντως ἐλεύθεροι ἔσεσθε.
John 8:33–36 (ESV 2001)	They answered him, "We are offspring of Abraham and have never been enslaved to anyone. How is it that you say, 'You will *become free'*?" Jesus answered them, "Truly, truly, I say to you, everyone who practices sin is a *slave to sin. [...] So if the Son sets you free*, you will be free indeed."

Paul's call for a united identity in Galatians 3:28 can be heard in the words of 1 Corinthians 12:13, where he writes of believers baptized into one spirit, whether Jew, Greek, slave or free: "εἴτε Ἰουδαῖοι εἴτε Ἕλληνες εἴτε δοῦλοι εἴτε ἐλεύθεροι"[12] (BNT). Paul employs the ethical sense of the binary and calls for a life of holiness resulting from freedom in Romans 6:22; he describes the believers as free from sin, and enslaved to God: "ἀπὸ τῆς ἁμαρτίας δουλωθέντες δὲ τῷ θεῷ"[13] (Romans 6:22, BNT). There are similarities in the uses of *ἐλευθερία* in Galatians and other books of the Pauline corpus yet it must be recognized that the letters were written for different purposes, for example Galatians, Ephesians and Philippians for groups of believers, whereas Timothy and Titus were written to individuals, therefore the sense of *ελευθερία* may have varied in its interpretation according to context.

Outside the Pauline corpus, there are occurrences of *ἐλευθερία* (freedom) and *δουλεία* (slavery) in the Gospels, 1 and 2 Peter, James and Revelation, which suggests that this dichotomy was already existent in the minds of believers in early Christian communities. In John 8, we read a conversation between Jesus and some Jewish leaders including the dichotomy (Table 6).

In John 8 *ἐλευθερία* is presented as a status that comes from God, opposed to *δουλεία*. Therefore we can see the use of this dichotomy beyond Galatians, both inside and outside the Pauline corpus.

2.3 Intertextual Occurrences

Some insight can be gained into this binary by looking at intertextual references to *ἐλευθερία* (freedom) and *δουλεία* (slavery). If, as Allen asserts, "the meanings we produce and find within language [...] are relational" (2000: 10) then it is valuable to look at other texts outside of the Biblical canon, where related or opposing senses of the keywords may occur. According to theologian Richard Longnecker, the call for unity despite economic and legal status in Galatians 3:28 stands in stark contrast

[12] MT: Jews or Greeks, slaves or free (1Corinthians 12:13b).

[13] MT: But, now you are set free from sin and bound to the service of God (Romans 6:22a).

to both "Jewish and Greek chauvinistic statements" (1990: 157). Paul writes, "There is neither Jew nor Gentile, neither slave nor free, nor is there male and female, for you are all one in Christ Jesus"(NIV) whereas the three berakot (blessings) at the beginning of the Jewish cycle of morning prayer state that: "Blessed be He [God] that He did not make me a Gentile; blessed be He that he did not make me a boor [i.e., an ignorant peasant or slave]; blessed be He that He did not make me a woman" (Longnecker 1990: 157). Moreover, Longnecker reminds us that "analogous expressions of "gratitude" appear in Greek writings of the time, for example, "that I was born a human being and not a beast, next, a man and not a woman, thirdly, a Greek and not a barbarian" (attributed to Thales and Socrates in Diogenes Laertius' Vitae Philosophorum 1.33). Paul employs concepts and patterns consistent in Hebrew and Greek traditions and presumably familiar to his readers, yet shifts the set of values attached to them.

2.4 Historical Sites of Interpretation

Re-mappings of *ἐλευθερία* (freedom) and *δουλεία* (slavery) are no doubt influenced by the interpretations that have been passed down through theologians and church history. Hans Dieter Betz, one of the principal Galatians' scholars in the twentieth century, describes *ἐλευθερία* as "the central theological concept which sums up the Christian's situation before God as well as in this world. It is the basic concept underlying Paul's argument throughout the letter" (1979: 255). As there is such a wealth of commentary and many different theologians who have written on Galatians, we are going to look at the manifestations of these views through the lens of two principal movements in which the understanding and interpretation of *ἐλευθερία* played a central role.

Firstly we turn to the sixteenth century Reformation movement. Martin Luther's interpretation of Galatians has no doubt had a major influence on the understanding of this dichotomy and its theological significance for Christians in Europe. Luther saw himself as a free man and identified the Medieval Church authorities as the Judaizers of his day.[14] He related *ἐλευθερία* and *δουλεία* to the religious realities of his time, yet focused on the necessity of salvation and freedom of the individual, with a particular focus on Galatians 2 verse 16, which he translated: "Doch weil wir wissen, daß der Mensch durch Werke des Gesetzes nicht gerecht wird, sondern durch den Glauben an Jesus Christus" (Galatians 2:16, Lutherbibel 1984).[15]

Four centuries later in the Roman Catholic Church, *ἐλευθερία* took on an important meaning in the Liberation Theology movement, which began in South

[14] Luther called Galatians his favourite epistle and even went so far as to call himself Eleutherius (Bainton 1955: 135).

[15] Translation (my own) of Luther's rendering of Galatians 2:16: Because we know that man is not made right by the works of the law but through belief in Jesus Christ.

America in the 1960s. This movement arose out of a reaction to social and political inequalities in South America. For Liberation theologians, *ἐλευθερία* signifies a release from corrupt political powers and social inequalities. There is a strong emphasis on the material and practical sense of *ἐλευθερία*. Gutierrez, one of the founding fathers of the movement, writes: "the coming kingdom [. . .] the peace and justice, the love and freedom [. . .] are not only private realities or internal attitudes, they are social realities, implying an historical liberation" (1988: 97). Other movements such as black liberation and feminist liberation then took their doctrine from this interpretation by Liberation theologians.

Although Lutherans and Liberation Theologians stemmed from different times and parts of the world, with Luther focusing on the individual's freedom, and Liberation theologians centered on social change and liberation of peoples, they both have one thing in common, according to theologian Rudolph Blank. He writes: "Unlike many First World theologians, both Liberation Theology and Lutheran Theology believe that the Scriptures are relevant to what is happening in the world today" (1994: 238).

Jewish scholars, on the other hand, have had a more conflictual relationship with Paul's letters and the concepts within them. According to Mark Nanos, this is because of the negative portrayal of Jews within the interpretative portrait of Paul built by Christians down the centuries.

> If one simply looks at what has been traditionally valued positively when Christians portray Paul, and at how that thinking is inextricably bound up with the way Judaism is negatively characterized in making these comparisons, one can readily see why Jews have traditionally understood Paul as an enemy and a danger to communal interests. (Nanos in Bird 2012: 160)

He lists common characterizations of the apostle Paul in popular Jewish imagination as "traitor, apostate, convert (and) deceiver" (Nanos in Bird: 160). How do translators for a Messianic Jewish audience combat this negative image of Paul, as they render concepts within his writings? How do other ideological markers come to light as we look at the rendering of *ἐλευθερία* in Bibles written specifically for Roman Catholic, Protestant and Messianic Jewish communities? As we take a comparative look at translations, how do concepts migrate between these cultural and confessional boundaries, and as Williams writes, "become actual alternatives in which the problem of contemporary belief and affiliation are contested?" (1983: 20).

3 Intercultural Map Analysis

The analysis of translations has been divided in terms of semantic nuances and therefore pragmatic implications of occurrences in translation. Following on from the thematic analysis of *ἐλευθερία* (freedom) and *δουλεία* (slavery) in the original Greek text, we now look at how new nuances are emphasized and various pragmatic implications come into play. The authorised versions (the old maps) are compared with new maps of freedom and slavery in Galatians, in terms of the characteristics

Table 7 Old maps of identity in Galatians 2:4

	Galatians 2:4
BNT	διὰ δὲ τοὺς παρεισάκτους ψευδαδέλφους, οἵτινες παρεισῆλθον κατασκοπῆσαι τὴν ἐλευθερίαν ἡμῶν ἣν ἔχομεν ἐν Χριστῷ Ἰησοῦ, ἵνα ἡμᾶς καταδουλώσουσιν,
	(MT: But through the secretly-brought-in false brothers, who slipped in to spy on our liberty, which we have in the Messiah Jesus, in order to enslave us.)
NRS (1989)	But because of false believers secretly brought in, who slipped in to spy on **t***he freedom we have* in Christ Jesus, so that they might enslave us.
RVA (1989)	a pesar de los falsos hermanos quienes se infiltraron secretamente para espiar *nuestra libertad que tenemos* en Cristo Jesús, a fin de reducirnos a esclavitud.
	(MT: Despite false brothers who secretly infiltrated to spy on our freedom that we have in Christ Jesus, with the goal of reducing us to slavery.)
LUT (1984)	Denn es hatten sich einige falsche Brüder mit eingedrängt und neben eingeschlichen, um *unsere Freiheit* auszukundschaften, *die wir in Christus Jesus haben*, und uns zu knechten.
	(MT: Because some false brothers had intruded and crept in alongside, in order to spy out our freedom, that we have in Christ, and to subjugate us.)
LSG (1910)	Et cela, à cause des faux frères qui s'étaient furtivement introduits et glissés parmi nous, pour épier *la liberté que nous avons* en Jésus -Christ, avec l'intention de nous asservir.
	(MT: And because of the false brothers who were introduced and snuck in among us, to spy out the liberty we have in Jesus Christ, with the intention of enslaving us.)

of freedom, its permanency, its authenticity, its relation to a Hebrew past and a Jewish present. Lastly, we will look at how the binary opposition has been presented as relevant or irrelevant to modern day divisions, through the strategies of the translator.

3.1 The Permanency of Freedom

The Permanency of Freedom Identity is the first consideration in the analysis of *ἐλευθερία (*freedom) and *δουλεία* (slavery). Translations may present *ἐλευθερία* as an inherent status of the believer, a possession which can be taken away, or a quantity that can increase and decrease. To look at this question we will examine translations of two verses: Galatians 2:4 and Galatians 5:1. Table 1 displays the rendering of Galatians 2:4 in the old maps (Table 7).

Each of these four traditional renderings of Galatians 2:4 suggest that ἐλευθερία stands for a possession, owned or gained by the believer. In correspondence with the Greek *ἣν ἔχομεν* (MT: we have), the translators have chosen phrases "the freedom we have" (NRS 1989), "nuestra libertad que tenemos" (RVA 1989), "unsere Freiheit [. . .], die wir [. . .] haben" (LUT 1984) and "la liberté que nous

avons" (LSG 1910).[16] Each translation above suggests that ἐλευθερία is something that is gained through Jesus Christ, and spied upon by opponents to the gospel.

However, the power of the opponents to take away this possession is interpreted and inferred differently by each translation. The NRS (1989) rendering "so that they *might enslave* us" implies an attempt by opponents to take away a liberty possessed by believers. However, the use of the modal "might" suggests this was an unsuccessful attempt at stealing *ἐλευθερία*. The Louis Segond (1910), "*avec l'intention* de nous asservir"[17] similarly creates some distance between the objectives of the opponents and their actual effect. This may have the theological implication that *ἐλευθερία* is something inherent, not easily snatched away by opponents. The Lutherbibel "um [...] uns zu knechten" (in order to subjugate/enslave us) is ambiguous about the ability of the opponents to take *ἐλευθερία* away from the believers, unlike the English modal "might" which suggests failure on the part of the opponents. *Reina Valera* (1989) "a fin de reducirnos a la esclavitud"[18] is similarly unclear on the success rate of the infiltrators.

With the use of "reducirnos" (MT: reduce us) in the RVA (1989), a vertical hierarchy is implied between *ἐλευθερία* and *δουλεία*. Those marked by *ἐλευθερία* stand above those who do not possess it. The translator widens the gap between the liberated and the enslaved and suggests the possibility of being reduced to a different status or moving up and down on a scale between *ἐλευθερία* and *δουλεία* (Table 8).

In the NRS (1989), LUT (1984) and LSG (1910), freedom, *Freiheit* and *liberté* are the end goals for the believer in Galatians 5:1, with the prepositions "*for* freedom", "*Zur* Freiheit" (to/for freedom) and "*pour* la liberté" (for the purpose fo freedom). *ἐλευθερία* is translated across these European contexts as an identity and status to be reached. The RVA (1989), however, does not translate the first clause in a literal sense, rather, brings the two clauses together, placing an emphasis on Christ. The translator renders the verse "estad, pues, firmes en la libertad *con que* Cristo nos hizo libres" (Galatians 5:1, RVA 1989). Here there is no end goal in the verse, as the preposition comes before Cristo. *Libertad* itself is the means by which Christ has given these new believers their new identity. In each of the four old maps there is a contrast between being under two different powers – either within the sphere of influence and firm in the

16

Translations from Corpus of Galatians 2:4b	Translation (my own) into English
"nuestra libertad que tenemos" (RVA 1989)	Our liberty that we have
"unsere Freiheit [...], die wir [...] haben" (LUT 1984)	Our freedom [...] that we [...] have
"la liberté que nous avons" (LSG 1910)	The liberty that we have

17 My Translation (MT) of LSG 1910: with the intention of enslaving us (Galatians 2:4).

18 MT of RVA: With the goal of reducing us to slavery (Galatians 2:4).

Table 8 Old maps of Galatians 5:1

BNT	Τῇ ἐλευθερίᾳ ἡμᾶς Χριστὸς ἠλευθέρωσεν · στήκετε οὖν καὶ μὴ πάλιν ζυγῷ δουλείας ἐνέχεσθε. (Gal 5:1 BNT)
NRS (1989)	For freedom Christ has set us free. Stand firm, therefore, and do not submit again to a yoke of slavery.
RVA (1989)	Estad, pues, firmes en la libertad con que Cristo nos hizo libres, y no os pongáis otra vez bajo el yugo de la esclavitud.
	(MT: Be, therefore, firm in the liberty with which Christ made us free, and do not put yourselves another time under the yoke of slavery.)
LUT (1984)	Zur Freiheit hat uns Christus befreit! So steht nun fest und laßt euch nicht wieder das Joch der Knechtschaft auflegen!
	(MT: For freedom Christ has freed us! Stand firm and do not let the yoke of slavery be laid upon you (plural, familiar) again.)
LSG (1910)	C'est pour la liberté que Christ nous a affranchis. Demeurez donc fermes, et ne vous laissez pas mettre de nouveau sous le joug de la servitude.
	(MT: It's for liberty that Christ has liberated us. Therefore remain firm, and do not let yourselves be placed again under the yoke of servitude.)

identity of ἐλευθερία, or under the dominion of δουλεία.[19] Each of these authorized versions translates quite literally the Greek "ζυγῷ δουλείας" (BNT) as "yoke of slavery"; "el yugo de la esclavitud"; "das Joch der Knechtschaft" and "le joug de la servitude."

In the authorized versions, there are two statuses and two definitive spheres of influence in both these verses. How do new re-mappings flesh out these traditional interpretations, in their various cultural and confessional contexts? (Table 9).

The *New Living Translation* (2007) and *Volxbibel* (2005) are both communicative translations, drawing from broadly evangelical traditions. Authorized versions in the respective English and German traditions opt for "spy on" and "auskundzuschaften" (to spy out/scout out) to describe the action the opponents took upon *ἐλευθερία* (freedom), with the consequences of their intended efforts not made clear. However, the NLT (2007) adds a clause and makes specific what perhaps is implicit in the Greek text. The NLT interprets the actions of the opponents in the first century and the implications for twenty-first century believers. The translators write "They sneaked in to spy on us and *take away the freedom* we have in Christ Jesus" (Galatians 2:4, NLT 2007). This implies that ἐλευθερία is not an everlasting identity whenever a believer enters into faith, but rather something that can be taken away from a believer, if they are not careful to keep it. Later, in Galatians 5:1 we read, in a similar line of interpretation: "Now make sure that you stay free." The VLX (2005), published in a similar time period to the NLT although

[19] Theologian Luke Timothy Johnson (2012 in Bird 2012: 68) conceptualizes this interpretation of the contrast between ἐλευθερία and δουλεία:

> Paul can speak of slavery or captivity to cosmic powers [. . .], to death [. . .], to the flesh [. . .], to sin [. . .], and to the law [. . .]. Rescue from such negative forces, in turn, can be expressed in terms of freedom [. . .], spirit [. . .] and life.

Table 9 New maps of identity

Galatians 2:4 NLT (2007)	Even that question came up only because of some so-called Christians there– false ones, really– who were secretly brought in. They sneaked in to spy on us and *take away the freedom* we have in Christ Jesus. They wanted to enslave us and force us to follow their Jewish regulations.
Galatians 5:1 NLT (2007)	So Christ has truly set us free. Now *make sure that you stay free,* and don't get tied up again in slavery to the law.
Galatians 2:4 VLX (2005)	Diese ganze Diskussion über diese "Beschneidung" wäre vermutlich gar nicht erst aufgekommen, wenn nicht ein paar Pseudos aufgekreuzt wären. Die schleimen sich zuerst mal richtig ein, um uns dann mit ihren Lehren *die Freiheit wieder zu klauen***,** die wir in Jesus gerade bekommen haben. Sie wollen nur, dass wir uns ihren religiösen Gesetzten total unterwerfen.
	(MT: This whole discussion about this "circumcision" would probably not have arisen, had a few pseudos not showed up. They really slimed their way in, to steal our freedom [that we just received in Jesus] with their teachings. They just wanted us to bow down completely to their religious rules.)
Galatians 5:1 VLX (2005)	Liebe Leute, Gottes Plan sieht für jeden von euch *eine totale Freiheit* vor! Trotzdem solltet ihr diese Freiheit nicht missbrauchen, indem ihr euch ätzenden Sachen ausliefert. Besser wäre, ihr nutzt sie, um euch gegenseitig zu lieben und zu respektieren.
	(MT: Dear people, God's plan provides total freedom for each of you! Nonetheless, you shouldn't abuse this freedom, by surrending to corrosive things. It would be better to use it to love and respect each other mutually.)

in a different cultural context, describes the intentions of Paul's opponents: "um uns dann mit ihren Lehren *die Freiheit wieder zu klauen*."[20] The translator, Martin Dreyer, uses a colloquial word for stealing "klauen", implying that ἐλευθερία represents a possession that can be taken away rather than a fixed identity. The permanency of the concept *ἐλευθερία* is questioned. In the description of the nature of *ἐλευθερία* obtained in Christ in Galatians 5:1, Dreyer renders part of the verse: "eine totale Freiheit", total freedom implies that partial freedom is a possibility and so, resonating with the RVA (1989) Galatians 2:4 "reducirnos a esclavitud" (reduce us to slavery) once again appears the idea of a scale from total freedom, to minimal freedom, to slavery.

3.2 The Authenticity of Freedom

Some translations suggest a scale of *ἐλευθερία* (freedom) with the possibility of sliding up and down, yet in other translations there is a clear binary opposition. The question here is the authenticity of *ἐλευθερία* rather than reaching a higher or lower

[20] MT: to steal our freedom (that we just received in Jesus) with their teachings.

Table 10 Galatians 5:1

	Galatians 5:1a
BFC (1996)	Le Christ nous a libérés pour que nous soyons *vraiment libres*. (MT: The Christ has liberated us so that we are truly free.)
LIT (1980)	Si le Christ nous a libérés, c'est pour que nous soyons *vraiment libres*. (MT: If Christ has liberated us, it's in order that we are truly free.)
LSG (1910)	C'est pour la liberté que Christ nous a affranchis. (MT: It's for liberty that Christ has emancipated us.)

level of a certain ontological state. Two of the French translations in the corpus present ἐλευθερία as an identity and there being clear boundaries marking the space inside and outside of true freedom (Table 10).

Both the BFC (1996) and the LIT (1980) move in a different interpretive direction from Louis Segond (1910), the more authorised version in French. Although the BFC (1996) and LIT (1980) translations come from differing confessional contexts (BFC is evangelical and LIT was authorized by the Catholic Church in Francophone countries) they both emphasize the authenticity of ἐλευθερία against the backdrop of either false ἐλευθερία or conventional δουλεία which are one in the same.[21] The concepts *verité* (truth) and *liberté* (liberty/freedom) are thereby connected in these interpretations of Galatians 5:1.

3.3 Freedom and the Hebrew Past

Given that the translations in the corpus come from various confessional contexts in European languages, it is interesting to examine how Paul relates this dichotomy to the Hebrew Scriptures and Jewish tradition. Nanos, a prominent Jewish scholar, describes what he considers a grave problem in the interpretative portrait of Paul, which has been passed down in tradition by Christians, and ends in a very negative perception of Judaism for Christians, while equally presenting Pauline theology as anti-Judaic:

> As Christians see it, Paul declared the "law-free gospel," [. . .] offering freedom from Judaism's mistaken focus on external rituals, on legalism, and seeking to earn God's favour. (In Bird 2012: 144)

How do we see this pattern interpreted in translations in these four linguistic contexts? How are *ἐλευθερία* (freedom) and *δουλεία* (slavery) connected to Judaism, and secondly, to the Hebrew Scriptures? (Table 11).

[21] This question of authenticity takes into account the overall narrative of Galatians, and Paul's opening statement, in which he displays his astonishment at the believers' willingness to turn away from the gospel to what he considers falsehood:

> Θαυμάζω ὅτι οὕτως ταχέως μετατίθεσθε ἀπὸ τοῦ καλέσαντος ὑμᾶς ἐν χάριτι [Χριστοῦ] εἰς ἕτερον εὐαγγέλιον, (Galatians 1:6 BNT)

Table 11 Old maps of Galatians 4:23

	Galatians 4:23
NRS (1989)	One, the child of *the slave*, was born according to the flesh; the other, the child of the *free woman*, was born through the promise.
RVA (1989)	Pero mientras que el de *la esclava* nació según la carne, el de *la libre* nació por medio de la promesa.
	(MT: But while the one of the slave was born according to the flesh, the one of the free was born by means of the promise.)
LUT (1984)	Aber der von der *Magd* ist nach dem Fleisch gezeugt worden, der von der *Freien* aber kraft der Verheißung.
	(MT: But the one of the maid was begotten according to the flesh, the one by the free was begotten by the power of the promise.)
LSG (1910)	Mais celui de *l'esclave* naquit selon la chair, et celui de *la femme libre* naquit en vertu de la promesse.
	(MT: But the [son of] the slave born according to the flesh, and the one of the free woman born under the promise.)

In Galatians 4, Paul introduces the story of Abraham from the Hebrew Scriptures. He makes the connection between the believers in the first century and prophecies in the Hebrew Scriptures, in Genesis. Paul uses the motif of *ἐλευθερία* and *δουλεία* to connect the Galatian community with Abrahamic promises. In Galatians 4:23 *ἐλευθερία* is equated with promise and *δουλεία* with flesh. There are two groups of people: those belonging to the blessings and promises of Abraham, marked by *ἐλευθερία* and those outside of them, marked by *δουλεία*.

In the Greek BNT there is a clear contrast between *παιδίσκης* (often translated as slave/servant) and *ἐλευθέρας* (the feminine nominative form of *ἐλευθερία*) in Galatians 4:23. Some translations emphasize the contrast between these two concepts and imply a hierarchy. The RVA (1989) opts for the rendering "el de la esclava [...]el de la libre" (MT: that of the slave [...] that of the free), and the German LUT (1984) in turn reads "der von der Magd [...] der von der Freien" (MT: that of the maid [...] that of the free), presenting a clear distinction between the sons of Abraham and consequently between those inside and outside of Abraham's inheritance. Yet the English NRS (1989) and French LSG (1910) create a further distinction and a vertical hierarchy in their divisions between the realm of the *ἐλευθερία* and *δουλεία*. The NRS (1989) makes a difference between "the child of the *slave*" and "the child of the *free woman*" whereas the Louis Segond renders the same two phrases "celui de *l'esclave* [...] celui de *la femme libre*" (that of the slave [...] that of the free woman). By inserting the idea of womanhood, the free becomes more embodied and human than the slave. Perhaps this represents a greater spiritualization of the story. The theological implications of this linguistic decision may be to see a greater distinction between a life lived in freedom and a life lived in slavery, between believers and those who do not believe. *ἐλευθερὶα* entails humanity and life in the renderings "la femme libre" (LSG 1910) and "the free woman" (NRS 1989) whereas the life marked by *δουλεία* and outside of Christ is lacking this life-giving element, with the simple renderings "la esclave" (LSG 1910) and "the slave" (NRS 1989). Another interpretation may be to connect the idea of slavery

Table 12 New maps of Galatians 4:23

	Galatians 4:23
CJB (1998)	The one by the slave woman [. . .] by the free woman
BFC (1996)	Le fils qu'il eut de la première [. . .] le fils qu'il eut de la seconde (MT: The son that he had to the first [. . .] the son he had to the second)
PDT (2005)	El hijo de la mujer esclava [. . .] el de la mujer libre (MT: The son of the slave woman [. . .] the (son) of the free woman)
SGB (2011)	El hijo de la esclava [. . .] el de la libre (MT: The son of the slave [. . .] the (son) of the free)

with immaturity and freedom with maturity "the free woman" and "la femme libre". These are just a few possible interpretations that may be drawn from subtle translation choices, although each reader comes with their own perspective to the text (Table 12).

In the new maps of this verse, the hierarchical distinction is not made between the female slave and the free female. In the SGB (2011) we read simply "la esclava" (the slave) and "la libre" (the free/liberated) whereas in the more communicative PDT (2005) the translators render the contrast "la mujer esclava [. . .] la mujer libre" (the slave woman [. . .] the free woman) thereby embodying both concepts and adding the human aspect to both parties. The BFC (1996) leaves out the mention of *ἐλευθερία* or *δουλεία* in this verse, perhaps in an attempt to modernise or domesticate and simply communicates the order in which both sons were born "la première" (the first (feminine)) and "la seconde" (the second (feminine)). This sample is indicative of the remainder of the corpus, where the contrast between *ἐλευθερία* and *δουλεία* is not widened by lexical choices which embody the concepts.

3.4 Freedom and the Jewish Present

After the story of Abraham and his wife is told, Paul goes on to connect this to his present day, to the city of Jerusalem, marked by *δουλεία* (Table 13).

νῦν Ἰερουσαλήμ, the present Jerusalem in verse 25 is contrasted with ἄνω Ἰερουσαλὴμ *ἐλευθέρα* (Gal 4:26 BNT), the free Jerusalem above, in verse 26. This could be interpreted as heavenly versus the earthly Jerusalem. In the NRS (1989) we read of the present Jerusalem "in slavery with her children" and the RVA (1989) corresponds quite closely to this with its rendering "esclava juntamente con sus hijos" (slave together with her children). However, the LUT (1984) with the word "Knechtschaft" and the LSG (1910) with the lexical choice "servitude" imply something between slavery and servanthood. *Sklaverei* or *esclavage* could have been chosen to render *δουλεία* in this context, yet the LUT (1984) and LSG (1910) bring out the more domestic element of the story, and the division between *ἐλευθερία* is not just as stark as in the NRS (1989) and RVA (1989) (Table 14).

Table 13 Hagar and Sarah

	Galatians 4:25
BNT	τὸ δὲ Ἁγὰρ Σινᾶ ὄρος ἐστὶν ἐν τῇ Ἀραβίᾳ• συστοιχεῖ δὲ τῇ νῦν Ἰερουσαλήμ, δουλεύει γὰρ μετὰ τῶν τέκνων αὐτῆς."
NRS (1989)	Now Hagar is Mount Sinai in Arabia and corresponds to the present Jerusalem, for she is *in slavery with her children*.
RVA (1989)	Porque Agar representa a Sinaí, montaña que está en Arabia y corresponde a la Jerusalén actual, la cual *es esclava juntamente con sus hijos*.
	(MT: Because Hagar represents Sinai, mountain which is in Arabia and corresponds to the present Jerusalem, that which is a slave with her children.)
LUT (1984)	Denn Hagar bedeutet den Berg Sinai in Arabien und ist ein Gleichnis für das jetzige Jerusalem, das *mit seinen Kindern in der Knechtschaft lebt*.
	(MT: Since Hagar signifies Mount Sinai in Arabia and is a symbol for the present Jerusalem, that lives with its children in bondage.)
LSG (1910)	Car Agar, c'est le mont Sinaï en Arabie, -et elle correspond à la Jérusalem actuelle, qui est *dans la servitude avec ses enfants*.
	(MT: Because Hagar is Mount Sinai in Arabia, and she corresponds to the present Jerusalem, that is in servitude with her children.)

Table 14 Hagar and Sarah re-mapped

	Galatians 4:25
CJB (1998)	Hagar is Mount Sinai in Arabia; she corresponds to the present Yerushalayim, for *she serves as a slave* along with her children.
NLT (2007)	And now Jerusalem is just like Mount Sinai in Arabia, because she and her children *live in slavery to the law*.
MSG (2005)	One is from Mount Sinai in Arabia. It corresponds with what is going on in Jerusalem- *a slave life, producing slaves as offspring*. This is the way of Hagar.
PDT (2005)	Agar representa el monte Sinaí en Arabia, y corresponde a la Jerusalén de hoy que *está en esclavitud* junto con el pueblo que ha dado a luz.
	(MT: Hagar represents Mount Sinai in Arabia and corresponds to the Jerusalem of today which is in slavery together with the people who gave birth.)
SGB (2011)	En efecto, Agar significa la montaña del SinaI, que está en Arabia, pero corresponde a la Jerusalén actual, pues *está sometida a esclavitud* junto con sus hijos.
	(MT: In effect, Hagar signifies Mount Sinai, which is in Arabia, but corresponds to the present Jerusalem, that is subjected to slavery together with her children.)
VLX (2005)	Hagar ist übrigens der arabische Name für den Berg Sinai. Er steht für unser heutiges Jerusalem, *für die Juden, die strikt an den Gesetzen kleben und nie wirklich frei werden, weil sie Jesus nicht vertrauen wollen*.
	(MT: Hagar is the Arabic name for Mount Sinai. It stands for our modern-day Jerusalem, for the Jews, who stick strictly to the laws and never become free, because they do not want to trust Jesus.)
VCE (2011)	Hagar is Mount Sinai in Arabia, and she stands for the Jerusalem we know now *she has lived in slavery along with her children*.

The Voice (VCE), by use of the perfect tense "she *has lived in slavery*" draws our attention to a condition of Jerusalem, and almost to a sense of pity or empathy for the present or earthly Jerusalem. The SGB (2011) similarly draws empathy from

the part of the reader with the use of a passive and the choice of verb *someter*, it reads "*está sometida* a esclavitud" (is submitted to slavery). The CJB (1998), which openly seeks to promote Jewish identity in the New Testament, renders the phrase "for she *serves as a slave*." The phrase attaches a noble idea to Jerusalem, serving faithfully. Judaism is therefore incorporated into the idea of ἐλευθερία, as δουλεία does not relate to a darkened way of thinking or evil opposed to God, but rather to a helpless and noble position, "serving as a slave".

The MSG (2005) and the PDT (2005) both suggest δουλεία as a way of life, with the respective phrases "what is going on in Jerusalem - *a slave life*, producing slaves as offspring" and "la actual ciudad de Jerusalén, donde todos sus habitantes son *esclavos de la ley*".[22] The PDT (2005) depicts the concepts of δουλεία, Judaism and the law on one side of the coin, with *ἐλευθερία* on the other.

The VLX (2005) is most explicit in its interpretation of Judaism (whether first century or twenty-first century Judaism) when rendering the keywords *ἐλευθερία* and *δουλεία*. The translator fleshes out the rather abstract idea of *νῦν Ἰερουσαλήμ* (Jerusalem now) for their modern day readers, and in doing so makes comment on Judaism. The translator adds to the quite literal "unser heutiges Jerusalem" (our Jerusalem of today) in the next explanatory clause "*für die Juden*" (for the Jews), thereby explaining that this verse about Jerusalem relates directly to Jewish identity. Instead of translating the phrase: *δουλεύει γὰρ μετὰ τῶν τέκνων αὐτῆς*. (Galatians 4:25 BNT), the VLX (2005) leaves out the reference to δουλεία, and instead interprets the implications of the enslavement. Martin Dreyer writes that the Jews are stuck to the law and will never experience *ἐλευθερία* fully because they do not trust Jesus: "strikt an den Gesetzen kleben und nie wirklich frei werden, weil sie Jesus nicht vertrauen wollen".[23] The Volxbibel explains that the old way of life, opposed to freedom is connected to the Jews, and inability or unwillingness to break away from the law. The VLX (2005), like the PDT (2005), and the NLT (2007) suggest a new dichotomy, of Christianity and Judaism, the first which brings freedom and the second which remains in slavery. In these interpretations there is no room for a Jewish freedom, or a continuation of Judaism together with the message and liberation of Christ, but rather there is a presentation of a break off from Judaism, likened to a release from captivity.[24] This is not to say that the VLX (2005) or NLT (2007) are inherently anti-Jewish translations, but just to show that theological interpretation of Paul vis-à-vis Judaism has implications on translation decisions, especially in relation to the dichotomy of *ἐλευθερία* and *δουλεία*.

Another simple contrast in interpretations relating to Judaism can be illustrated in the first mention of ἐλευθερία in the letter, in Galatians 2:4. Table 16 shows the difference between renderings in the *Complete Jewish Bible* (1998) and the *New Living Translation* (2007) (Table 15).

---

[22] My Translation (MT) of PDT (2005): the current city of Jerusalem, where all the inhabitants are slaves to the law.

[23] See Table 14 for translation.

[24] This is what Mark Nanos describes as the interpretative portrait of Paul often propagated by Christians in recent times, in which "every strength attributed to Christianity is compared to a negative one attributed to Judaism" (Nanos in Bird 2012: 144).

Table 15 Jewish freedom or slavery

	Galatians 2:4
NLT (2007)	They sneaked in to spy on us and take away the freedom we have in Christ Jesus. They wanted to *enslave us* and *force us to follow their Jewish regulations*.
CJB (1998)	They came in surreptitiously to spy out the *freedom* we have *in the Messiah Yeshua*, so that they might enslave us.

Table 16 neither δοῦλος nor ἐλεύθερος

	Galatians 3:28
NRS (1989)	There is no longer Jew or Greek, there is no longer slave or free, there is no longer male and female; for all of you are one in Christ Jesus.
RVA (1989)	Ya no hay judío ni griego, no hay esclavo ni libre, no hay varón ni mujer; porque todos vosotros sois uno en Cristo Jesús.
	(MT: There is no longer Jew nor Greek, slave nor free, there is neither man nor woman; because all of you (informal, plural) are one in Christ Jesus.)
LUT (1984)	Hier ist nicht Jude noch Grieche, hier ist nicht Sklave noch Freier, hier ist nicht Mann noch Frau; denn ihr seid allesamt einer in Christus Jesus.
	(MT: Here is neither Jew nor Greek, here is neither slave nor free, here is neither man nor woman; because you (informal, plural) are all one together in Christ Jesus.)
LSG (1910)	Il n’y a plus ni Juif ni Grec, il n’y a plus ni esclave ni libre, il n’y a plus ni homme ni femme; car tous vous êtes un en Jésus -Christ.
	(MT: There is no longer Jew nor Greek, there is no longer slave nor free, there is no longer man nor woman; for all of you are one in Jesus Christ.)

By opting for “*Messiah* Yeshua”, the translator of the CJB (1998) David H. Stern connects freedom to the Hebrew Scriptures and to Judaism, whereas the NLT (2007) translators, in contrast, connect the act of enslaving to “Jewish regulations”, following on from the line of interpretation mentioned above.

3.5 Freedom and Slavery: An Ancient or Modern Division?

Now we will turn to a few examples of domestication or foreignization of *ἐλευθερία* (freedom) and *δουλεία* (slavery) and look at how this affects interpretation. This is taken from Schleiermacher’s conceptualization of domesticating and foreignization, wherein domestication involves bringing the text to the reader with more familiar features to the target audience and foreignizing involves keeping the distance between the reader and the text, retaining the foreign or alien features. Previous examples related to translation the relationship between Christians and Jews, whereas this example and the various translation choices relate to the relationship between believers regardless of background (Table 16).

This verse describes an end in division between believers, and a new unified identity in Christ, despite social, legal, religious and gender differences. Yet

Table 17 New maps of Galatians 3:28

	Galatians 3:28
CJB (1998)	*There is neither* Jew nor Gentile, neither *slave nor freeman*, neither male nor female; for in union with the Messiah Yeshua, you are all one.
SGB (2011)	*No hay* judío y griego, *esclavo y libre*, hombre y mujer, porque todos vosotros sois uno en Cristo Jesús.
	(MT: There is no Jew and Greek, slave and free, man and woman, because all of you (plural, familiar) are one in Christ Jesus.)
PDT (2005)	*Todos son iguales* en Jesucristo, *no importa si* son judíos o no, *esclavos o libres*, hombres o mujeres.
	(MT: All are equal in Jesus Christ, it doesn't matter if they are Jews or not, slaves of free, men or women.)
EIN (1980)	*Es gibt nicht mehr* Juden und Griechen, *nicht Sklaven und Freie*, nicht Mann und Frau; denn ihr alle seid «einer» in Christus Jesus.
	(MT: There are no longer Jews and Greek, no slaves and free, no man and woman; because you (plural, familiar) are all "one" in Christ Jesus.)
VLX (2005)	Ein paar Sachen existieren für uns einfach nicht mehr: zum Beispiel *der Unterschied* zwischen Jude und Nichtjude, *zwischen Arbeiter und Unternehmer* oder zwischen Mann und Frau. Wir sind jetzt alle eins, weil wir zu Jesus Christus gehören.
	(MT: A few things simply don't exist for us any more: for example the difference between Jew and non-Jew, between worker and employer or between man and woman. We are now all one, because we belong to Jesus.)

ἐλευθερία (freedom) and *δουλεία* (slavery) are not legal divisions that the Western readers of Scripture face on an everyday basis in the twenty-first century. Does this division have significance for readers of the Bible today? In these authorised translations, the concepts *ἐλευθερία* and *δουλεία* and the division between them remains quite abstract and therefore they may easily be spiritualized by readers (Table 17).

The new maps of Galatians 3:28 have mostly followed on from the old cartography, as the plain idea of there not being division between those marked by *ἐλευθερία* and those marked by *δουλεία* continues, with a foreignizing and distancing effect on the reader. There is one exception in the corpus, in the VLX (2005) translation, where the translator applies a domesticating strategy to the two concepts and brings the idea of division into a twenty-first century European context of division. The division is between "Unternehmer" and "Arbeiter", where *Unternehmer* refers to an employer, businessman or contractor and *Arbeiter* usually to a person who is employed and often carries out physical labour. The difference between *ἐλευθερία* and *δουλεία*, therefore, like in the original context, displays a social rather than a religious difference in the phrase *οὐκ ἔνι δοῦλος οὐδὲ ἐλεύθερος* (Gal 3:28 BNT). Although *δοῦλος* and *Arbeiter* have less semantic correspondence than *δοῦλος* and *Sklave*, which the EIN (1980) uses, it could be argued that in translating this dichotomy as *Unternehme*r and *Arbeiter*, the VLX (2005) translation serves a similar pragmatic function to the original text. This domesticating strategy has its own theological implications. In translating the dichotomy as *Unternehmer* and *Arbeiter*, Martin Dreyer compels the reader to let their faith

Table 18 Galatians 5:13

	Galatians 5:13
BNT	ὑμεῖς γὰρ ἐπ᾽ ἐλευθερίᾳ ἐκλήθητε, ἀδελφοί · μόνον μὴ τὴν ἐλευθερίαν εἰς ἀφορμὴν τῇ σαρκί, ἀλλὰ διὰ τῆς ἀγάπης δουλεύετε ἀλλήλοις. (Gal 5:13 BNT)
RVA (1989)	Vosotros fuisteis llamados a la libertad, hermanos; solamente que no uséis la libertad como pretexto para la carnalidad. Más bien, servíos los unos a los otros por medio del amor.
	(MT: You [plural, familiar] are called to liberty, brothers; only do not use your liberty as a pretext for carnality. Better than that, serve one another by means of love.)
PDT (2005)	Hermanos, Dios los ha llamado para ser libres. Pero no permitan que la libertad sea una excusa para complacer sus deseos perversos. Mejor ayúdense los unos a los otros siempre con amor.
	(MT: Brothers and sisters, God has called you [plural, formal] to be free. But do not allow the liberty to be an excuse to indulge your perverse desires. Better help one another out, always with love.)
SGB (2011)	Pues, vosotros, hermanos, habéis sido llamados a la libertad; ahora bien, no utilicéis la libertad como estímulo para la carne; al contrario, sed esclavos unos de otros por amor.
	(MT: Then, you [plural, familiar], brothers, have been called to liberty; now well, do no use the liberty as stimulus for the flesh; on the contrary, be slaves one to another in love.)

have an impact on social division within the church and society. Therefore, there is an echo, in a very different confessional context (the Volxbibel is from an evangelical tradition), of the doctrines of Liberation theology, which called for a consideration of social problems and class distinctions.

The PDT (2005), although it translates *ἐλευθερία* (freedom) and *δουλεία* (slavery) in a typical fashion, as "esclavos o libres" (slaves or free people), leans in the direction of this consideration of social issues and divisions within the church, with the opening phrase "*Todos son iguales* en Jesucristo, *no importa si*" (all are equal in Jesus Christ, it doesn't matter if), therefore laying more emphasis on the necessity to look beyond divisions and find unity among believers actively.

A domesticating strategy in translation may have a similar theological implication in Galatians 5:13. In the three Spanish translations below, we see the renderings of *δουλεία* (Table 18).

In this context, Paul turns the *ἐλευθερία* and *δουλεία* dichotomy on its head, calling the believers to be enslaved to oneanother in love. The SGB (2011) follows quite closely in semantic correspondence with the rendering "sed esclavos." The RVA (1989) and PDT (2005), domesticate the concept *δουλεία*, calling on believers to "servíos los unos a los otros por medio del amor" (RVA), to serve oneanother through love, and respectively "ayúdense los unos a los otros siempre con amor", helping each other out, through love. The domestication strategy brings the modern reader closer to the text and emphasizes the ethical nature of *ἐλευθερία* as a way of life, and of living in community with other believers.

4 Conclusion

In conclusion, this paper demonstrates the value of the study of sacred texts within the field of Pragmatics. The readers expectations, the context of the concepts and their implications all interconnect to form meanings for the reader, and often differ in emphasis according to group or affiliation.

By examining the background to the dichotomy *ἐλευθερία* (freedom) and *δουλεία* (slavery), its journey through different interpretations, and the manifestations of ideology in translation, it is clear that the use of keywords reflects ideological affiliations, and that relationships between keywords are in fact embedded within real-life relationships. Certain renderings of *ἐλευθερία* (freedom) and *δουλεία* (slavery) may be either indicative or prescriptive of relationships between confessional groups and relationships between believers in community. Interpretations of Galatians affect translations and equally translations have an impact on interpretations of the texts, as most readers of the Bible have little access to ancient Greek. New dichotomies are formed, others are abolished and new connections between utterances and concepts are formed. An interpretive portrait of Paul vis-à-vis Judaism, passed down for centuries in Europe, may be propagated in a translation of Galatians 2:4 "they wanted to enslave us and force us to follow their Jewish regulations" (NLT 2007), yet this same interpretive portrait may be challenged by another translation that appropriates Jewish identity in the English text "the freedom we have in the Messiah Yeshua" (Galatians 2:4, CJB).

The theological implications of *ἐλευθερία* (freedom) and *δουλεία* (slavery) for the individual reader in the twenty-first century are interpreted and re-written by translators with varying degrees of emphasis on the ethical or ontological aspects of *ἐλευθερία* and *δουλεία*. The PDT (2005) translation of Galatians 5:13 "ayúdense los unos a los otros siempre con amor" (helping each other out in love), may lead to pragmatic implications for the reader in their understanding of the concept of freedom and its implications within community, while another translation may focus more on the individual.

The VLX (2005) applies a domesticating strategy to bring *ἐλευθερία* and *δουλεία* into a twenty-first century context of class distinction with the lexical choices "Arbeiter" and "Unternehmer", and thereby implies social liberation, in contrast to every other translation in the corpus, both old and new, which keep the metaphor of slavery. This translation may not seem closely correspondent to the Source Text, yet perhaps it carries over the pragmatic function of the text more clearly than many of its counterparts within the corpus.

The old maps (the authorised translations) in English, French, German and Spanish appear at first to be less interpretive and often stay closely to semantic correspondence of the Greek text. Yet in Galatians 4:23 we see that the NRS (1989) and the LSG (1910), two of the authorised translations widen the gap between *ἐλευθερία* and *δουλεία* with their lexical choices, rendering the distinction as "the slave [. . .] the free woman" (NRS 1989) and "l'esclave [. . .] la femme libre", thereby creating a dehumanized impression of the enslaved in comparison with the free. In contrast, all of the new maps within the corpus do not take this interpretive leap.

From the analysis of both the background to these concepts and their journey through Bible translations, it is evident that the choices of the translator of sacred texts are not merely linguistic choices, but rather they are often rooted in theological decisions. The analysis of keywords and connections between keywords within a biblical book in translation provides an excellent tool to discover patterns and differences across confessional and cultural boundaries, to look at the journey of concepts across contemporary Christian discourse, and finally to understand better the dialogue between groups and affiliations.

References

Aland, K, et al. (Eds.). (2001). *Novum Testamentum Graece (BNT)* (27th ed.). Stuttgart: Deutsche Bibelgesellschaft. (BibleWorks. v.9)

Allen, G. (2000). *Intertextuality*. London: Routledge.

Bainton, R. (1955). *Lutero* (R. L. Ayala Torales, Trans.). Buenos Aires: Editorial Sudamericana.

Bell, D. (2010). *A comparative analysis of formal shifts in English bible translation*. Saarbruecken: Lambert Academic Publishing.

Betz, H. D. (1979). *Galatians: A commentary on Paul's letter to the churches in Galatia*. Minneapolis: Fortress Press.

Bird, M. (Ed.). (2012). *Four views on the Apostle Paul*. Grand Rapids: Zondervan.

Blank, R. H. (1994). Six theses concerning freedom in Christ and liberation: Liberation in Galatians, Luther, and liberation theology. *Concordia Journal, 20*(3), 236–260.

Blumczyński, P. (2006). *Doctrine in translation. The doctrine of the trinity and modern English versions of the New Testament*. Lask: LEKSEM.

Blumczyński, P. (2007). Pneuma in Galatians: A touchstone in doctrinal orthodoxy. In *Anglica Wratislaviensia XLV*. Wroclaw: University of Wroclaw.

Chomsky, N. (1987). Language in a psychological setting. *Tokyo: Sophia Linguistica, 22*, 1–73.

Die Bibel: Nach der Übersetzung Martin Luthers. (1984) Stuttgart: Deutsche Bibelgesellschaft.

Die Volxbibel 2.0: Neues Testament (2nd ed.). (2005). Köln: Volxbibel-Verlag.

Einheitsübersetzung der Heiligen Schrift (EIN). (1980). Stuttgart: Katholische Bibelanstalt GmbH. (BibleWorks, v.9)

Gutierrez, G. (1988). *A theology of Liberation* (2nd Ed., Translated from Spanish by Inda, C., & Eagleson, J.). New York: Orbis Books.

Holy Bible, New Living Translation (2nd ed.). (2007). Carol Stream: Tyndale House Publishers.

Johnson, L. T. (2012). The Paul of the letters: A Catholic perspective. In M. Bird (Ed.), *Four views on the Apostle Paul* (pp. 65–96). Grand Rapids: Zondervan.

Kahl, B. (2010). *Galatians re-imagined: Reading with the eyes of the vanquished*. Minneapolis: Fortress Press.

Kecskes, I. (2010). Situation-bound utterances as pragmatic acts. *Journal of Pragmatics, 42*, 2889–2897.

La Bible de la liturgie. (1980). Paris: Association Épiscopale Liturgique pour les pays Francophones.

La Bible en francais courant. (1996). Paris: Société biblique francaise.

La Bible: Louis Segond. (1910). Paris: Société Biblique Française.

La Biblia: La Palabra de Dios para Todos (PDT). (2005). Texas: Centro Mundial de Traducción de La Biblia.

La Santa Biblia Reina-Valera Actualizada (RVA). (1989). n.p.: de la Casa Bautista de Publicaciones (BibleWorks, v.9.).

Longnecker, R. N. (Ed.). (1990). *Word biblical commentary: Galatians*. Dallas: Word Books.

Nanos, M. (2012). A Jewish view. In M. Bird (Ed.), *Four views on the Apostle Paul*. Grand Rapids: Zondervan.

New Revised Standard Version Bible (NRS). (1989). New York: Division of Christian Education of the National Council of the Churches of Christ in the United States of America. (BibleWorks, v.9)

Peterson, E. (2005). *The Message: The New Testament in contemporary English*. Colorado Springs: NavPress.

Sagrada Biblia: versión oficial de la Conferencia Episcopal Española. Edición popular. (2011). Madrid: Biblioteca de Autores Cristianos.

Société Biblique Française. (1996). *Le nouveau testament illustré en français courant*. Villiers-le-Bel: Brodard & Taupin.

Stern, D. H. (1998). *Complete Jewish Bible: An English version of the Tanakh (Old Testament) and the B'rit Hadashah (New Testament)*. Clarksville/Jerusalem: Jewish New Testament Publications, Inc.

The Holy Bible, English Standard Version (ESV). (2007). Wheaton: Crossway Bibles. (BibleWorks, v.9)

The Holy Bible: New International Version (NIV). (2011). Colorado Springs: International Bible Society. (BibleWorks, v.9)

The Voice New Testament. (2011). Nashville: Ecclesia Bible Society.

Wansbrough, H. (Ed.). (1985). *New Jerusalem Bible*. New York: Doubleday. BibleWorks, v.9.

Wierzbicka, A. (2003). *Cross-cultural pragmatics: The semantics of human interaction* (2nd ed.). Berlin: Mouton de Gruyter.

Wierzbicka, A. (2004). Jewish cultural scripts and interpretation of the Bible. *Journal of Pragmatics, 36*(3), 575–599.

Williams, R. (1983). *Keywords: A vocabulary of culture and society*. London: Fontana Press.

Wilson, D., & Sperber, D. (2004). Relevance theory. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 607–632). Oxford: Blackwell.

How to Make People Feel Good When Wishing Hell: Golden Dawn and National Front Discourse, Emotions and Argumentation

Fabienne H. Baider and Maria Constantinou

Abstract Drawing on Kristeva's thesis that nationalism is a pervasive discourse of exclusion based on "a defensive hatred" where "the cult of origins easily backslides to a persecuting hatred" (Kristeva 1993), the main argument of this chapter is to demonstrate that extreme right parties such as the National Front in France and Golden Dawn in Greece are constructing a different ethos, having the same far-right political aims. In particular, the analysis of their discursive practices for constructing differently their political arguments (defensive resentment vs persecuting hatred) are based on different emotions as well (contempt vs. pride) while appealing to the same argumentative stance (the cults of origins for instance). Indeed, if they both deliver the same nationalistic discourse of "reciprocal exclusion", Marine Le Pen's persona, discourse and behaviour are based, as will be argued here, on a "defensive resentment", focusing on the Self via the emotion of *pride*. On the other hand, Golden Dawn's leader, Nikos Michaloliakos, uses gestures, speeches and symbols, which encourage a virulent "persecuting hatred", focusing on the despised Other via the emotion of *contempt*. Focusing on the metaphors and linguistic symbolisms used to embody these emotions, we will show using van Dijk's theory of ideology in discourse and corpus linguistic methodology, that these tropes are conducive to legitimating, triggering and perpetuating social practices of exclusion of some specific communities as well as symbolic or physical violence against the same communities.

Keywords Extreme right discourse • National Front • Golden Dawn • Metaphors • Symbolism • Pride • Contempt • Specificity

F.H. Baider (✉) • M. Constantinou
Department of French Studies and Modern Languages, University of Cyprus,
Po Box 20537, CY 1678 Nicosia, Cyprus
e-mail: fabienne@ucy.ac.cy

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2,
DOI 10.1007/978-3-319-06007-1_9, © Springer International Publishing Switzerland 2014

1 Nationalistic Discourse: Defensive Resentment and Persecuting Hatred

Even though many studies have been conducted in the fields of pragmatics and intercultural studies aiming to bridge cultural differences (Angouri 2010) and promoting cross-cultural awareness, the opposite trend, protectionism and nationalism, is looming in the field of politics. Due to anxiety over recession, high levels of unemployment and immigration, far-right wing parties have considerably grown and gained breeding ground. As a matter of fact, the themes of *borders* and *national identity* were the focus of many 2012 French and Greek presidential elections speeches (Sarkozy and Le Pen political campaigns for France and Golden Dawn for Greece). This chapter aims to demonstrate that extreme right parties such as the National Front in France and Golden Dawn in Greece are constructing a different ethos, through similar but complementary discursive practices (defensive resentment vs. persecuting hatred), while appealing to the same argumentative power of emotions. Marine Le Pen's persona, discourse and behaviour are based, as will be argued in this study, on a "defensive resentment", focusing on the Self via the emotion of *pride*. On the other hand, Golden Dawn's leader, Nikos Michaloliakos, uses gestures, speeches and symbols which encourage a virulent "persecuting hatred", focusing on the despised Other via the emotion of *contempt*. Based on both written (articles about the leaders, political agenda and ideology of the National Front and Golden Dawn) and oral data (speeches by and interviews with Parties' leaders and members), this chapter focuses on discursive (metaphors) and semiotic strategies employed to embody these emotions. The theoretical framework consists of an integration of Rhetorical Psychology (Billig 1978 for instance) and Critical Discourse Analysis (henceforth CDA) (van Dijk 1995 et passim) approaches; our methodology includes corpus linguistic tools such as AntConc and TermoStat and takes into account statistical parameters such as frequency and specificity.

2 From Lexical Semantics to Discourse Analysis

In June 2012, Golden Dawn, the extreme right party, won 6.7 % of the vote and 18 seats in the Hellenic Parliament. Ten years before in France in 2002 J.-M Le Pen, candidate for the Front national, beat the socialist candidate L. Jospin in the first round during the presidential elections and gained nearly 17 % of votes. In 2012, Marine Le Pen, his daughter and the new party representative, was a major threat for Sarkozy during the first round. Even though he was the former French president, Sarkozy had to take on Marine Le Pen's arguments for immigration and national identity during his mandate in order to win the 2012 elections.

Nationalism was looming in both instances. Both parties are also known for their extreme right tendencies as are their (ex-)leaders[1] for their xenophobic statements. To examine the discursive strategies which embody emotions used as political arguments, the present study bridges academic research in lexical semantics on hatred (and its associated emotions) and discourse analysis of extreme right parties, as language is considered, in this chapter, to be at the core of psychological constructs.

2.1 Discourse Analysis, Ideological Stance

Our analysis aims to identify the mechanisms of racist rhetoric in the mediated discourse of the two political groups as well as the views of their respective leaders. Our aim is not to generalize the reactions and opinions of the actors involved but to illustrate how a covert fascist discourse (Marine Le Pen's stance) can contain identifiable rhetoric mechanisms, which have the same illocutionary impact as an overt Nazi discourse (Golden Dawn's speeches for instance). In the limited space of this chapter, we have selected data which could best illustrate the expression of these mechanisms because of the vocabulary used, the organization of the text, and the links between Nazi ideas.

Leaders of both parties under investigation in this paper, the National Front and Golden Dawn, speak, write and understand (and make believe) reality from a specific ideological position in order to defend or legitimate their social practices to be enacted by their followers (van Dijk 1995). This is achieved in terms of a strong polarization: Us (= good, innocent) and Them (= evil, guilty) (van Dijk 2006b). Our analysis based on CDA as developed by van Dijk (1995 and 2006c in particular) focuses on political argumentation, and more precisely, on the argumentative use of *emotions*, which will be perceived in this chapter as being a social, cognitive and discursive phenomenon (van Dijk 2006b). Political use of emotions is a *social* phenomenon inasmuch as it involves relationships of power between the politicians and their audience; it is a *cognitive* phenomenon since emotions are viewed here as being first thoughts to be provoked (Wierzbicka 1992); it is also a *discursive* phenomenon since emotions are conveyed here through talk and text. Therefore, our study relies first on empirical reality, the actual socio-political context, which is then mediated by speech. Moreover, the study of discursive structures will seek to explain how emotions are produced and reproduced in order to legitimate or create a specific socio-political agenda.

According to van Dijk (2006a), the overall strategy of *ideological positioning* consists in the polarized presentation/action of Us as positive and that of Them as negative as already indicated. This is achieved by emphasizing *Our* good things,

[1] Jean-Marie Le Pen (JMLP) has retired and Nikos Michaloliakos having been charged for crime, his wife Ms Zaroulia has taken the leadership position.

and *Their* bad things, and de-emphasizing *Our* bad things, and *Their* good things. Although this seems a simplistic view of ideology, this positioning is achieved with an extremely skilled use of (among other elements) specific rhetorical structures such as *repetition*, *euphemisms, comparisons*, *metaphors*, *metonymies,* etc. Adopting van Dijk's framework, Potvin et al. (2004) have identified the following strategy to differentiate and categorize individuals in the two groups (US-Them) in nationalistic discourse:

– for the Self and the Other: generalization of certain traits or individual behaviour group while ignoring the reality faced by the Other;
– for the Other: inferiorization, essentialization or/and demonization, biologization so much that the discourse will call for the expulsion of the Other;
– for the Self: self-victimization and political legitimation;
– for the Nation: Construction of catastrophic scenarios.

Taking into consideration these categories as guidelines for this research, we are focusing on how emotions embodied in conceptual metaphors and political symbolism are used as indirect justifications for inferiorizing, excluding, or exploiting a specific community, and/or valorizing the Self, thereby achieving the rhetoric aims listed above.

2.2 Emotions Creating Subjectivity

We perceive emotions as being a special type of a socially shared script, which manifests itself in both language and behaviour of a given social group (Kitayama and Masuda 1995). Our semantic and pragmatic study is, therefore, anchored within a *socio-cognitivist framework* (Kecskes 2011, 2012; Wierzbicka 1988 *et passim*, Lazarus 1991; Scherer 2000) and *Conceptual Metaphor Theory* (Lakoff and Johnson 1980, 1999; Kövecses 1986 *et passim*) at the micro-grammar of the sentence. From a cognitivist perspective and an appraisal standpoint, emotions are the result of the assessment of a particular situation (Lazarus 1997 *et passim*, Frijda 2007). As Smith and Lazarus (1990) explain, the role of emotion in psychology is fundamental:

> Given the central position that we cede to emotions in our personal lives and the prominence of emotion in literary studies of the human condition, one might expect emotion to serve as a central, organizing construct in scientific psychology, and especially in a psychology of personality.

Given also the importance of psychology in argumentation and especially in political argumentation, it is no surprise that emotions play a major role in persuasion. As such, emotions – especially fear and anger – have been studied as political arguments in political rhetoric (Marcus 2000; Holmes 2004; Mio et al. 2005; Frijda 2007; Groenendyk 2011) even though the sociology of emotion had been 'neglectful of the political' (Holmes 2004, 210).

From a socio-cultural point of view, social relationships produce and organize emotions (Kemper 2001 *inter alios*), in other terms, we are actually 'doing' emotions according to some rules and norms (Averill 1980). However, from a post-modern stance, Holmes considers that emotions *make* rather than simply emerge from subjects and the relations between them: the reciprocal dynamic of emotions being productive of social relationships and being produced by these social relationships imply their performative dimension (Holmes *ibidem*, 211–212). Indeed, emotions motivate behaviours and actions (the "appetitive aspect" of the emotion, cf. Lyons 1985); for instance, anger motivates people to redress injustice (Holmes *ibidem*, 224). Hence, the importance of 'brainwashing' (cognitive element) to induce the targeted emotions (feeling element) that, in turn, will motivate the targeted behaviour (social dimension).

The appeal to emotions helps not only to construct an 'emotional identity' but also contributes to reinforcing an ethos of power inherent in the populist discourse for constructing the Self in relation to the Nation as we will see in Le Pen's discourse. It also reorients one self and another person to each other (Frye 1983) as we will see in the discourse of Golden Dawn. For instance, the repulsion felt towards the Other when being angry will imply a lack of respect for this Other. This feeling of repulsion is not static; it is "part of communicative processes that produce subjects in relation to others" (Holmes *ibidem*, 213). Emotions are then situated and embodied in activity that inhibits or makes possible ongoing relationships with Others (Cornell 1995). Therefore, in order to create irremediable tensions within a community, extremist discourses have recourse to more dysphoric emotions such as hatred, contempt, and hubris.

Previous studies (Billig 1978; Savage 2007; Kohl 2011) on Nazi discourse agree that the main targeted emotion of Nazi discourse is hatred. Hatred remains a rare topic of research in linguistics. Limited previous research on hatred, namely in criminology, identified it as the emotion having a cognitively based motivation since it is an evaluative and negative judgement of the target of the emotion; it is thus motivated by beliefs, attitudes and thoughts (Kahan and Nussbaum 1996; Rosebury 2003). However, Perry's study (2006) focusing on the outcomes of the emotion (hate crime and hate speech) concludes that violence triggered by hatred "relies heavily on the history and persistence of relations of advantage and disadvantage" (2006, 135). Therefore, any feeling of hatred has to draw on available cultural sources for meaning and legitimacy as well as to show advantage and its opposite. From the point of view of lexical semantics, and according to a socio-cognitive theory on language (Kecskes 2008, 2011), prominent semantic features in lexical units, which are derived from social experiences and specific historical encounters, will shift and new meanings will be assumed for old words. These are anchored in the everyday and repetitive use of the language and are the basis for conceptual associations, available to each speaker of that speech community.

Drawing on previous work on the lexical semantics of hatred (Baider 2012, 2013), which was based on different associations, identified in oral and written data, we concluded that:

- the emotion of **resentment**, found especially in interviews, is also identified prominently in extreme right discourses. Indeed, hatred is described as an emotion that lingers (one develops a feeling of hatred, as one informant said) and nourishes itself on resentment.
- the presence of the word **contempt** is prominent in journalistic, personal and literary discourse. It presupposes the negative focus on another person previously mentioned; previous research emphasized that the construction of an 'evil other' is a sine qua non component in extreme right discourses;
- scorn could explain, then, the desire for **vengeance, revenge** and **the will to harm** or even destroy the other present in all types of discourse investigated with the salient feature of revenge; as for the Nazi discourse, the need to awake, a need for revenge have also been identified; this entails taking action, such as the denunciation of neighbours as is the case with what had happened during the Nazi occupation;
- the emotion of **anger including its** extreme version **rage, was the most common one,** in all our data, with the emotion of **fear** in its extreme meaning, **terror**.

All these emotions are found at the core of speeches, behaviours and beliefs of parties that call themselves 'nationalists' such as the National Front and Golden Dawn.

2.3 Political Argumentation, Metaphors and Emotions

Convincing speakers, from all political spectrums, have recourse to both logos and pathos to construct their discursive ethos: linguistic strategies are then intermingled with emotion utterances, evaluative modalisation, punctuation marks, capital letters, etc. (Amossy 2008; Plantin 2011). However, emotions in a speech will not be designated as such. On the contrary, they will be indexically referred to in the speech, with no clear emotion term present in the discourse. They will be inferred in the description of physiological emotional states (*red-in-the brow* indexes shame for French), or metaphors (such as *parasites* to refer to unwanted people) as far as they are identified in the discourse through stereotypes (Plantin 2011).

Many previous studies have identified the impact of using metaphors as well as symbolic language, labels and slogans in political persuasion and charismatic leadership (Willner 1984; Shamir et al. 1993; Shamir et al. 1998; Conger 1991; Mio 1997). From a cognitive viewpoint, metaphor is defined as a cross-domain mapping (Lakoff 1993, 203) whereby "one experiential domain is partially 'mapped', i.e. projected, onto a different experiential domain, so that the second domain is partially understood in terms of the first one" (Barcelona 2000, 3). In other words, conceptual metaphors provide us with unfamiliar ways of conceptualizing familiar things, and familiar ways of conceptualizing unfamiliar things (Lakoff and Johnson 1980). For instance, the conceptual metaphor "ARGUMENT IS WAR",

allows us to view argument as a battle to be won (as linguistic expressions such as *she won the argument* testify); hence the social practice to have a political debate as a 'face à face'. This metaphor shapes the way we think of argument and the way *we go about* arguing.

Previous works on the third Reich discourse have shown how much metaphors and symbolisms create a powerful short-cut reasoning in order to convince people to adopt a point of view (Billig 1978; Burke 1984; Chilton 2005; Charteris-Black 2004, 2005; Musolff 2007, 2008, *inter alios*). The metaphors embodying the emotion of contempt, one of the most frequently associated emotions with hatred – as identified in our research – along with symbolism are used to express and trigger its quasi-opposite, the emotion *pride* (as well as hubris i.e. overestimation of one's own competence, accomplishments or capabilities). These two emotions form the basis as we claim, for the political arguments of the extremist discourses studied in this chapter.

3 Emotions for an Out-Group Aversion and an In-Group Cohesion

This section identifies the main discursive characteristics of the Third Reich discourse, which is, by definition, a discourse of hatred. In particular, we examine how contempt and pride are fundamental to the Nazi propaganda. This review of the literature on Nazi discourse (Hitler and Goebbels' speeches mainly) identifies the triangular schema of this particular political propaganda:

- the main conceptual metaphors relating to the Other and the Self (linguistic expression);
- the divisive emotions (fear, anger, resentment and frustration towards the Other) and cohesive emotions (pride, hope for the Self) targeted by this recurrent linguistic usage (cognition);
- the effects of these emotions i.e. a group cohesion (uniforms, chant, groups, etc.) and an out-group aversion (ghettos, immigration laws, etc.) (social actions).

These discursive mechanisms, which often coalesce, are then called "socio-cognitive" (Windisch 1978, cited in Potvin et al. 2004), because they refer to both the reasoning and the thought of individuals as well as their political, ideological, cultural stance (among others).

3.1 Metaphors, Emotions and the Third Reich Discourse

Most scholars working on the discourse of the Third Reich agree that the otherness of the target is emphasized (Savage 2007; Musolff 2008); the Other is described or

Table 1 Emotions to be created towards the out-group

Negative emotions such as anger (resentment, frustration) and fear (apprehension) both leading to take action	
ANGER- RESENTMENT FEAR	Frustration and resentment are the result of the feeling of injustice, almost always present in the discourse on hatred;
	To that effect, the Nazis accused the Jewish community of being at the root of the loss of the 1st World War; they called Hebrews the Jewish community, with a view to recalling the death of Jesus Christ. (*They have killed before* (Goebbels);
FEAR – AVERSION	The minorities and the Jewish community were described as an obvious threat who wanted "to exterminate whatever is German" (Goebbels) (Jewish conspiracy);
	They were *dangerous*: *Even though they are a minority, they have great power* (Goebbels cited in Kohl 2011);
	Evil: the *bearer of international culture destroying chaos;*
	Not human: their otherness is emphasized (Savage 2007)

Table 2 Emotion to be created inside the in-group

The self embodies what is good, positive	
PRIDE:	To avoid paralysis with too much fear the strength of Germany is also evoked to incite to action. *Aryan shines, divine spark, genius, knowledge, allow man to rise and be the master of the universe* (Adolph Hitler cited in Kohl, 11)

referred to as subhuman, parasite, fungus, etc. (the conceptual metaphor of NON HUMAN or DEHUMANIZATION). The aim is to create a homogeneous and evil out-group, so that people will not feel guilty of harming or killing them since they are dangerous, subhumans, parasites, or cause or spread diseases. Table 1 shows which and how emotions were expressed and conceptualized in the Nazi discourse according to previous studies. The Tables 1 and 2 are based on studies carried out by Kohl (2011) on the Nazi discourse.

Indeed, this table corroborates lexical semantics research on hatred, mainly 'made out of' anger because of resentment due to injustice – real or imagined – which always leads to a desire for *revenge* (Baider 2013). Fear is also important to create the emergency that this survival instinct triggers.

The result of the blending of these emotions is a desire for an act of revenge and vengeance, which leads to the emotion of contempt against the targeted groups, on which we will be focusing in our analysis of Golden Dawn's discourse. These emotions will then be embodied in metaphors to galvanize the audience; for instance, the parasite metaphor, superbly studied by Chilton in the Nazi discourse, allows a formidable cognitive shortcut of all the emotions above in the blending of the Jew and the parasite:

> The first occurrence of *he* is associated not just with 'the Jew', but with a blended concept: Jew- parasite, or some such, and that the successive clauses predicate actions and properties that are metaphorically isomorphic with the actions and properties of biological parasites (Chilton 2005, 39).

Musolff (2008, 5) further explains that "the blend parasite = Jew thus built up and reinforced grammatically within the text with the effect that its chances of becoming a memorable meme are maximized". Once the cognitive shortcut is established ("the meme is remembered") the entailments of this conceptual metaphor can be "filled in and elaborated further within the disease and medicine frames and their specific aetiologies" (Musolff *ibidem*, 5).

At the same time common characteristics have to be constructed in speech so that the in-group members have common goals to defend (Billig 1978), i.e. a strong need to conform to obtain the superiority feeling and the material security provided by the in-group membership. Therefore, an ingroup cohesion is also sought through the emotions of pride and hope. Pride entails satisfaction with the Self when having or achieving something special that the Self or other people admire as the proper sense of one's own value. It can though encourage a feeling of superiority shown by somebody who believes, often unjustifiably, that he or she is better than others in its form of hubris as mentioned before. Indeed, previous studies on this emotion distinguish two facets (Tracy et al. 2010, 163–164; Hart and Matsuba 2007):

– the positive pride of oneself for good work, good deed and good self-esteem;
– the negative pride (*hubris*) seen as a non-deserved self-satisfaction which leads, in turn, to a feeling of (ill-deserved) superiority towards the others.

This second facet of pride, as will be argued here, is the targeted emotion of nationalistic discourse, since, in turn, hubris leads to contempt of the others. When discussing nationalistic discourse Potvin et al. (2004) have noted the expression and exploitation of the emotion of *pride*, an emotion also commented by Marlin during the Dreyfus affair in 1898:

> There are obvious differences in these cases, but *the hatreds and fears evoked were similar*. A form of *nationalist or patriotic pride* was at the centre of both. (. . .) *Once a mob is excited and has a rallying focal point,* raising skeptical doubts has no impact. The skeptic is simply lumped in with the enemy. (Marlin 1998 cited in Potvin, our italics).

Indeed, focusing the crowd on one emotion (such as *contempt* as we have just discussed), helps the in-group to be polarized.

When observing the main emotions analyzed in previous studies on Nazi speeches, we notice what van Dijk calls ideological positioning, which is indeed the main basis of such discourses a very clear division between good and evil.

3.2 Methodology

To appraise the ideological dimension of the texts under study, we will restrict our analysis to the lexical marking (collocations, specificities and frequencies of lexical units) based on previous studies (Bednarek 2008; Banks 2007; van Dijk 2006a; Tutin et al. 2006; Sinclair 2004; Blumenthal 2002).

This will allow us to analyze the construction of negative polarity and intensity in the texts. Quantitatively, the collocational properties of important lexical units

previously mentioned such as *immigrants* will show the value and belief system via the analysis of these lexical units used as well as the recurrent metaphors, which target the emotions described above. Statistics have been drawn in order to identify the most frequent words as well as the most frequent collocates of target words such as immigrants, parasites (metanastes, parasita) and the specificities of their speech (National Front). For the statistics dimension AntConc[2] was used since it supports also the Greek language, which is not the case for other softwares such as TermoStat.[3] From a qualitative perspective, AntConc allowed us to study paragraphs where the NP of each candidate appears in the article (KWIC abstract). The software TermoStat (Lesage et al. 1993; Drouin 2003) was used to examine both the most *frequent* and the most *specific* lexical categories for each corpus under study. The *specificity* parameter is a dimension taken into account in corpus linguistics which is defined as "the calculation of the difference between the relative frequencies of the linguistic items in the analysis and reference corpora"[4]: it is then the difference of frequency of lexical items in our data comparing them to the 'reference corpus' (28 millions of words) from *Le Monde* (2002).

If the quantitative analysis focuses on the most frequent words, collocates and specificities of our data, and in particular on the lexical units known to trigger the targeted emotions, the qualitative aspect will be assessed partly with KWIC abstracts for understanding the contextual environment of the lexical units and collocates under investigation. A comparison between the speeches held by the two parties (Golden Dawn and National Front) will then be attempted in order to corroborate our argument that these parties put into place different strategies for the same goal.

3.3 Data

Our study has been based on both written and oral data, in order to draw the most frequent words, collocates and specificities in both parties' discourses.

For the collection of Golden Dawn data, we worked on small context-specific corpora. We restricted our study only to Golden Dawn' official site, given that what is of importance here is how the Other is perceived by Golden Dawn. In this context, for our written corpora (of about 130,000[1] words[5]) in addition to articles by members of Golden Dawn appearing on their official website, we have also taken into consideration letters and messages by their partisans posted on the same site. The total number of articles by members of Golden Dawn amounts to 165 and that of their followers' letters accounts for 46. Our choice to include such letters and

[2] http://www.antlab.sci.waseda.ac.jp/software.html

[3] http://termostat.ling.umontreal.ca/doc_termostat/doc_termostat_en.html

[4] TermoStat (http://linguistech.ca/TermoStat_E_TUTCERTT_I_PartI)

[5] For the benefits of using small corpora in the analysis of pragmatic features see Vaughan and Clancy 2013.

messages[6] is not arbitrary, since they do not differ from the 'official' articles in terms of their ideological content, since they fully express and endorse the party's agenda, ideological positioning and anti-immigration stance. The first data were initially collected in May 2013 and completed over the period of October-December 2013 and January 2014 to cover the most recent events and dramatic changes in the movement after Pavlos Fyssas's murder (September 2013) and consequently the arrest of the leader and other members of Golden Dawn. The oral data include speeches by the Leader and other members of the party; more precisely, we have included the leader's speeches after Golden Dawn's electoral victory of the first and the second round in parliamentary elections in 2012, the 3rd speech delivered on 16th June 2013 to celebrate the continuing rise of Golden Dawn. We have also included the speech of Golden Dawn's president's wife, E. Zaroulia delivered on 18th October 2012, and Kasidiaris' speech, the spokesperson for Golden Dawn.

As for the National Front, our data include speeches by Jean-Marie Le Pen especially when conceding defeat in 2008 and 2012 to contrast with the speeches given by Marine Le Pen during the 2012 campaign and interviews to the press (see bibliography). Each interview or speech amounts to 1,500 – 2,000 words. In total for the oral data we have 20, 000 words. As for the written data, we had collected all the articles referring to the candidates during the 2012 presidential campaign (1, 220. 000 words roughly), with for each candidate 20 articles (cf. Baider and Jacquey 2014 for methodology and corpus description). For Marine Le Pen the corpus amounts to 160, 000 words. For Sarkozy, who adopted some of the National Front arguments during the campaign, the corpus amounts to 201,400 words. For each politician, the corpus is divided into three parts (roughly 65,000 words for each) which correspond to special timing of the presidential campaign: March – July 2011; September – December 2011; April – May 2012. We built the corpus from six different daily newspapers (regional and national newspapers, representing a wide political spectrum from left to right).

4 Analysis

The analysis focuses on the rhetorical force of the emotions necessary to motivate the public to act or at least to adopt the advocated ideology.

4.1 A Persecuting Contempt: The Parasite Metaphor in GD Speech

From a socio-cognitivist approach (Koselak 2005), the concept of 'contempt' is an internalised negative emotion towards the object of the emotion which is followed

[6] The column is called *Η Φωνή του Λαού* "The People's voice" (I foni tu lau).

by a certain behaviour. The object of the emotion of contempt is contested in a radical fashion as long as the value judgement is always and obviously negative: the object is classified durably in the category of "worthless, not good entity". Contempt is then unacceptable when applied to a community (contempt for students, people, Jews, women, etc.). As Koselak explains, the vagueness of the object when referring to a certain community means that because they belong to the group named women, Jews or Blacks, this entity will be judged as being bad, weak, worthless, etc.; under these circumstances, the subject of the emotion gives him/herself a tremendous "ethical" discriminatory power over the other: to place in contempt someone is to place the entity below oneself, which implies, of course, that oneself is above (Koselak 2005).

Linguistic uses from the Nazi past, which openly express contempt, are easily found in discourse held by Golden Dawn. To be more precise, E. Zaroulia (wife of the ex-GD president and the new president of GD)[7] said during a 2012 Greek parliament session that "immigrants are sub-humans", "invade Greece" and "spread all sorts of diseases". The expression *subhuman* used by Ms Zaroulia underlying the parasite metaphor is directly taken from Hitler and Goebbles' speeches (Untermenschen). Golden Dawn's doctrine also proposes that "only men and women of Greek descent and consciousness should have full political rights." Golden Dawn's ex-leader, N. Michaloliakos (20. 05. 2013), had publicly and repeatedly denied the Holocaust:

> What about Auschwitz?...I have not been to Auschwitz ...What happened [there]? ... I have read many books questioning the number of six million Jews. This exaggeration was fabricated... There were no ovens. This is a lie... There were no gas chambers either.

Several emotions have been the focus of journalistic comments about Golden Dawn's activities; fear on the part of the potential victims of the 'packs of activists', hatred on the part of the Golden Dawn's partisans, despair and desperation for the causes of joining Golden Dawn's electorate. For instance, the statement by Golden Dawn's leader after their electoral triumph "The time for fear has come for those who betrayed this homeland. We are coming." is highly revealing of their thoughts and the emotions they intend to provoke to the public. Fear, anger and hatred are intermingled with *contempt* since immigrants are inferiorized as well as Greeks (media, government, etc.) who do not adhere to their ideology are demonized or even "otherized" as we show in the entailments of the Parasite metaphor.

The construction of the Other as an evil entity in relation to the Nation can be characterized by the Nazi prevalent metaphor of parasite, very recurrent in Golden Dawn's discourse.

According to Oxford dictionary, a parasite is "an organism which lives in or on another organism (its host) and benefits by deriving nutrients at the other's

[7] Ms Zaroulia is paradoxically a member of the Greek parliament delegation to the Council of Europe's Committee on Equality and Non-Discrimination in Strasbourg. Her mandate was reconducted in 2013 despite protests from anti-racism organizations. (see http://www.politics.ie/forum/eu/197748-neo-nazi-golden-dawn-member-sit-eu-committee-non-discrimination.html)

expense". However, in order to draw all submappings concerning the parasite metaphor in relation to Golden Dawn's social practices and to what we have found in our written and oral data, this general definition should be completed with further scientific details. In fact, biology or botany dictionaries or other sources[8] inform us about other damaging and/or pathogenic aspects of parasites, since they can cause harm or diseases to their host.

Although they are generally much smaller than their hosts, they can affect health, performance and carcass quality in a serious way. Parasites may have an indirect life cycle. This means that a vector or intermediate host is required in order for the parasites life cycle to be completed. With this type of life cycle, one infected animal cannot infect another animal, it is the intermediate host which spreads the parasite. In this context, intermediate hosts are Troika, the media, the coalition government, the leftists and especially the strongest rival of Golden Dawn, Syrisa, who are accused of having caused or maintaining this dramatic situation based on a Zionist conspiracy plan. To denounce these factors, in addition to the word 'parasites', in their discourse we have found other animalistic terms such as cockroaches, rats, rodents, caterpillars, subhumans etc. To protect the nation, it is urgent to get rid of the intermediate hosts, so that the parasites (immigrants) cannot any longer be spread.

Each characteristic of parasites mentioned above corresponds to an entailment which is identified in Golden Dawn's speech and, in turn, explains social practices employed or advocated by that party. The variety of entailments displayed by this conceptual model are also associated with the emotions that they are likely to trigger to the public. Given the limited space here, we will give only one or two examples for each entailment.[9]

Entailment 1. A parasite invades a body to which it does not belong: the nation is the host body.

(1) The phenomenon of illicit (illegal) immigration has become a major problem [...] it is a guided invasion under an undeclared war that threatens to destroy our national sovereignty.[10]

Words in the example denote how immigrants are seen. Invasion, which is believed to be well planned and organized, implies an undeclared war, a constant

[8] Sources used are mainly from the Biology on line dictionary and Wikipedia: http://en.wikipedia.org/wiki/Host_(biology) http://girldogandscience.wordpress.com/2012/08/15/caterpillars-and-parasitoids/

[9] To facilitate the non Greek-speaking reader, we have placed the English translations in the text and the original texts and their transcription in footnotes.

[10] Το φαινόμενο της παράνομης (λαθραίας) μετανάστευσης, έχει αναδειχθεί σε μείζον πρόβλημα [...] πρόκειται για καθοδηγούμενη εισβολή στα πλαίσια ενός ακήρυκτου πολέμου, που απειλεί να καταλύσει την εθνική μας κυριαρχία [...] [To fenomeno tis paranomis (laθreas) metanastefsis, eχi anaδiχθi se mizon provlima [...] prokite jia kaθodiγumeni isvoli sta plesia enos akiriktu polemu, pu apili na katalisi tin eθniki mas kiriarχia].

threat and danger against their national survival. This entailment triggers and/or intensifies mainly fear, anger and hatred, while diminishing the feelings of mercy or guilt that Greeks may feel towards them. It is a kind of devictimisation. Riots, violent acts to exterminate or force them to leave the country constituting social 'good' practices followed by Golden Dawn or their partisans, are also related to this entailment.

Entailment 2. Parasites are generally much smaller than their hosts. In this context, immigrants or other undesirable are considered to be smaller in value or inferior to the nation.

(2) This fire started a year ago and has many *decadent parasites* to "burn" on its road [...][11]

The use of the derogatory term *decadent* aims obviously to elicit the emotion of contempt. Parasites do not deserve to stay on their land; they should be mercilessly burned, exterminated, since they are inferior to the Greek nation and have no place in their lives. A social practice related to this entailment is the creation of Blood banks, reserved only to Greeks and actually to those who fully embrace their ideology.

Entailment 3. The conceptualization of immigrants and other undesirable as parasites leads also to the insistence on differences in culture, ideology, behaviours (sexual orientation) etc., since a parasite is, by definition, a foreign body. For instance, the leftists are called *red parasites*, while immigrants are labelled as *third world invaders*.

(3) They devise unlawful laws in order not to offend *illegal third-world invaders* with sentences that reach up to disenfranchisement and we do not protect our country and ourselves from anti-Greek traitors [...].[12]

(4) Even though these fields exist, their extraction only serves the interests of the Greek capital, as claimed by the *red parasites* of the Nation.[13]

The use of such negatively charged modifiers triggers the emotion of contempt.

[11] Αυτή η πυρκαγιά έχει ξεκινήσει εδώ και έναν χρόνο και έχει πολλά παρακμιακά παράσιτα να «κάψει» στην διαδρομή της. [Afti i pirkajia eχi ksekinisi aδo ke enan χronon ke eχi polla parakmiaka parasita na kapsi sti διaδromi tis].

[12] Επινοούν λαθρολάγνους νόμους μην τυχόν και προσβάλουμε τους παράνομους τριτοκοσμικούς εισβολείς με ποινές που φτάνουν μέχρι και στέρηση πολιτικών δικαιωμάτων και εμείς δεν προστατεύουμε την πατρίδα μας και τον εαυτό μας από ανθέλληνες προδότες [...]. [epinoun laθrolaγnus nomus min tiχon ke prosvalume tus paranomus tritokosmikus isvolis me pines pu ftanun meχri ke sterisi politikon δikeomaton ke emis δen prostatevume tin patriδa mas ke ton eafto mas apo anθellines proδotes].

[13] Ακόμη όμως κι αν αυτά τα κοιτάσματα υπάρχουν, η εξόρυξή τους δεν αποτελεί παρά εξυπηρέτηση των συμφερόντων του ελληνικού κεφαλαίου, ισχυρίζονται τα ερυθρά παράσιτα του Έθνους. [Akomi omos ki an afta ta kitasmata iparχun i eksoriksi tus δen apoteli para eksipiretisi ton simferondon tu ellinku kefaleu, isχirizonde ta eriθra parasita tu eθnus].

Entailment 4. Parasites spread easily and very quickly so do immigrants. Despite their tininess parasites rule the world (see Jews as being viewed by Nazi) and any feeling of guilt or culpability should disappear, since they jeopardize the survival of the nation by increasing dangerously fast. Out of fear to be ruled by them, Golden Dawn is for their radical extermination:

(5) The situation is deteriorating day by day, since the number of illegal immigrants *is increasing with breakneck pace* on a daily basis and no serious measures are being taken to *prevent* and *combat* illegal immigration.[14]

Entailment 5. Parasites use the blood and/or get food by eating the host's digested food, depriving the host of nutrients to the point of destruction.[15] In this sense, immigrants are accused of taking or wasting the social product without giving anything back to the nation, since they are useless, cheaters, have no education, etc.:

(6) 'I am 25 years old and I will not allow any longer this plunder! *THESE PARASITES* do not pay taxes as we do! They do not send *their children to school and get social benefits* as having many children! They *do not serve the Greek army*! And of course *they do not pay their bills*![16]

(7) People have woken up… *their blood has been sucked* by the partners of our disaster […][17]

Because of taking the blood or getting too many nutrients, parasites cause decomposition of the host body; this may result in the host's death. For instance, the economic death that Greece has been going through over the last years or other social problems related to the invasion/overflow of immigrants or other undesirable, such as unemployment, violence, etc. This entailment is mostly related to the negative emotions of despair, anger and hatred, which, being accumulated, lead to the emotion of contempt.

[14] Η κατάσταση μέρα με την ημέρα επιδεινώνεται, δεδομένου πως ο όγκος των λαθρομεταναστών αυξάνει με ιλιγγιώδη ρυθμό καθημερινά και δε λαμβάνονται σοβαρά μέτρα για την πρόληψη και καταστολή της λαθρομετανάστευσης. [I katastasi mera me tin imera epiδinonete, δeδomenu pos o ongos ton laθrometanaston afksani me ilingioδi riθmo kaθimerina ke δen lamvanonde sovara metra jia tin prolipsi ke katastoli tis laθrometanastefsi]

[15] Hawkins referring to the iconography of the Human body points out that in popular understanding parasites 'maintain life within their own bodies by sucking life-sustaining nutrients out of some other body' (Hawkins 2001, 46).

[16] Είμαι 25 χρονών και δεν θα επιτρέψω άλλο αυτή την λεηλασία!! ΑΥΤΑ ΤΑ ΠΑΡΑΣΙΤΑ δεν πληρώνουν φόρους όπως εμείς! Δεν στέλνουν τα παιδιά τους στο σχολείο και παίρνουν επιδόματα πολυτέκνων! Δεν υπηρετούν τον ελληνικό στρατό! Και δεν πληρώνουν φυσικά τους λογαριασμούς τους! [Ime 25 χronon ke δen θa epitrepso allo afti tin leilasia!! AFTA TA PARASITA δen plironoun forus opos emis! Δen stelnun ta peδja tus sto sχolio ke pernun epiδomata politeknon! Δen ipiretun ton elliniko strato! Ke δen plironun fisika tus loγarjasmus tus!].

[17] Ο λαός έχει ξυπνήσει, του έχουν ρουφήξει το αίμα οι συνέταιροι της καταστροφής μας. [O laos eχi ksipnisi, tu eχun rufiksi to ema i sineteri tis katastrofis mas].

Entailment 6. Some parasites are pathogens; that is, they live within the host and cause diseases or injuries. Foreigners are believed to spread diseases by contaminating the nation, as also Zaroulia claimed in one of her speeches against immigrants. Therefore, the host body must avoid contact and defend itself.

(8) [. . .] Greece needs to start to behave with the ethos of a Lord and not that of a servant regarding national issues, to cleanse his Land, having been watered with his Blood, of *any racial and social parasites* that *defile* it and get *a robust* national leadership in a Fair and Steel State.[18]

This need for self-protection may be interpreted in the suggestion of making of new laws, in the creation of ghettos to avoid contamination etc. or in other social practices through the enactment and embodiment of the emotions previously studied (violence against them, feeling of resentment/contempt towards them, etc.) or the foundation and management of a Counseling Health and Care Center by Golden Dawn and its members. In this context, Golden Dawn acts as a medicine or a healer to cure and save the nation.

Entailment 7. A parasite implies also an intermediate host body, which also carries diseases and contaminates the body of the nation. Troika, the coalition government (and in particular Samaras), the leftists and especially their strongest rival, Syrisa are seen as being the intermediate hosts, since they are not called simply as parasites but they are labelled as cockroaches, caterpillars and rodents. Based on the alleged Zionist conspiracy, those intermediate hosts do not only aim to materially destroy the Greek nation (by facilitating or not controlling the flow of immigrants in the country) but also to cause its spiritual decomposition and death:

(9) Our fight rages, fire of freedom will be burning their anti-Hellenic cockroaches and rodents [. . .] *artificial spiritual famine* imposed on us *by International Parasites* will be *eradicated* [. . .] (Fear and Anger).[19]

(10) Nowadays' human *caterpillars* have withered the spiritual *full of freshness tree of our race*[20][. . .].

[18] [. . .] να αρχίσει η Ελλάδα να φέρεται με την ηθική του Κυρίου και όχι του δούλου στα εθνικά της θέματα, να καθαρίσει η ποτισμένη με το Αίμα του Γη από τα φυλετικά και κοινωνικά παράσιτα που τη μιαίνουν.

και να αποκτήσει μια στιβαρή Εθνική Ηγεσία σε ένα Δίκαιο και Ατσάλινο Κράτος [. . .] [na arχisi i Ellaδa na ferete me tin iθiki tu Kiriu ke oχi tu dulu sta eθnika tis θemata, na kaθarisi i potismeni me to ema tu Ji apo ta filetika ke kinonika parasita pu ti mienoun ke na apoktisi mia stivari eθniki ijesia se ena δikeo ke Atsalino Kratos].

[19] Ο αγώνας φουντώνει, η πυρκαϊά της ελευθερίας θα κάψει τις ανθελληνικές κατσαρίδες και τα τρωκτικά. [. . .] θα σβήσει ο τεχνητός πνευματικός λιμός που μας επέβαλαν τα Διεθνή Παράσιτα. [O aγonas fundoni, i pirkajia tis elefθerias θa kapsi tis anθellinikes katsariδes ke ta troktika [. . .] θa svisi o teχnitos pnevmatikos limos pu mas epevalan ta dieθni parasita.]

[20] οι σημερινές ανθρωποκάμπιες μαράζωσαν το πνευματικό ολόδροσο δέντρο της φυλής μας [. . .] [i simerines anθropokambjes marazosan to pnevmatiko oloδroso δendro tis filis mas].

Harking back to the glorification of origins, Golden Dawn aims to further intensify that persecuting hatred. For instance, the full of freshness tree in (10) symbolizes the nation and its spiritual life throughout the centuries. This tree has its roots in their ancestors, mainly from Ancient Greece to the Greek revolution, and the dangerous parasites and other intermediate host bodies harbour on the nation's body while causing its decomposition.

Entailment 8. Parasites are associated with dirt; hence the need to *purify* the nation in Golden Dawn's speech and the message to send home non-Greeks since they are perceived as being garbage, trash *(σκουπίδια)*. To face the problem of dirt, cleanliness is very important. Golden Dawn acting as "a national disinfectant" (cf. Papaioannou 2013), aims to radically get the stench out of the country, by implementing violent purifying strategies to get rid of foreigners viewed as garbage.

(11) Cleanse the country of the dirty illegal immigrants who have devastated Greece.[21]

Dirt is directly associated with crime:

(12) GOLDEN DAWN will bring back the country to the righteous path and restore criminal wrongdoings against OUR GREEKS which have been committed by rodents and worms of the underworld [...][22]

Social practices deriving from this metaphor include cleansing operations such as visiting streets and flea markets and violently getting rid of immigrants or visiting and cleansing houses occupied by immigrants to the benefit of their Greek owners. Such practices come to fill in the gaps in terms of state protection and correct the wrongs that the socialists, the justice and the police forces have created.

Immigrants' unlawful and criminal status is emphasized by the obvious preference for the use of the composed words *lathrometanastefsi* "illegal immigration" (29 occurrences) *lathrometanastis/es* (213 occurrences) "illegal immigrant(s)", which are more negatively charged than the less frequently used words *metanastefsi* (4 occurrences) "(im)migration" *metanastes (im)migrants (*34 occurrences).[23]

[21] Ξεβρωμίστε τον τόπο από τους βρωμιάρηδες λαθρομετανάστες που έχουν διαλύσει την Ελλάδα. [Ksevromiste ton topo apo tus vromiariδes laθrometanastes pou eχun δialisi tin Ellaδa.]

[22] Η ΧΡΥΣΗ ΑΥΓΗ θα επαναφέρει στον δίκαιο δρόμο τις εγκληματικές αδικίες εις βάρος των ΕΛΛΗΝΩΝ ΜΑΣ από τα τρωκτικά και σκουλήκια του υποκόσμου [...] [I XRISI AVJI θa epanaferi ston δikeo δromo tis englimatikes adikies is varos ton ELLINON MAS apo ta troktika kai skulikja tu ipokosmu.]

[23] It is relevant to note that according to our data the term *metanastes* in the Golden Dawn discourse is rarely used to refer to immigrants. The occurrences we have identified are part of reported speeches (by other people in power) which are employed to ironize and denounce the political agenda on immigrants adopted by the Government or other political parties or authorities. When the term *metanastes* is to refer to migrants in Greece, it appears along with the adjective *illegal* to stress their unlawful status or it is even used alone in ironical contexts such "poor economic migrants". The term is also used to refer to Greek migrants in other countries, or in other more general contexts such as International Migrants' Day.

A more derogatory term identified in our corpus is the neologism *lathroisvolis*/*lathroisvoli* "illegal invaders/invasion" (15 occurrences) which also serves to magnify immigrants' evil entity. As for the noun *immigration*, it collocates mostly with *illegal* and *uncontrollable*.

Our data reveal that GD discourse about immigrants revolves around three main thematic axes:

(a) criminality which comprises violence, thefts, rapes, human trafficking etc.
(b) rights and social benefits for immigrants at the expense of Greek people including their legalization and social problems that their stay in the country may cause (such as unemployment)
(c) incivility of immigrants emphasizing racial, religious, and cultural otherness.

Thematic axes	Number of tokens	Lexical units	Emotions triggered	Main entailments associated with
Criminality-danger/unlawfulness	287 tokens related to unlawfulness 89 tokens related to criminality	Illegal immigrants, illegal invaders, illegality, violence/violent acts, kill, murder, offenders, rapes, criminals, (commit horrible) crimes, criminality, trafficking, smugglers, smuggling, attacks, prison, disaster, danger black money, terrorists, terrify, Interpol, wanted, arrested, very severe charges, unscrupulous, predatory attacks	Fear, terror, anger, hatred	1, 2
Preferential treatment at the expense of Greek people	69 tokens	Attribution of benefits to immigrants, (human) rights for immigrants, policy *in favour of* illegal immigrants, parasite, legalization, public hospitals, free healthcare, for free illegal immigrants in public hospitals, unemployment (for Greeks), disaster, facilities, immigrants are accommodated, live luxuriously, luxurious comfortable houses, (The government) cherishes and gives food to immigrants etc./ protect immigrants, increase, large wound etc.	Despair, envy, anger and hatred	4, 5

(continued)

Thematic axes	Number of tokens	Lexical units	Emotions triggered	Main entailments associated with
Incivility and cultural/ religious/ racial otherness	52 tokens	Third world, (Afro-Asian) Muslims, trash, subhumans, contaminate, Roma, Pakistans, islamists, uncivilised, dirty, diseases, contaminate, minaret, mufti, idolaters, pagans, heathen, insults etc.	Hatred and contempt	2, 3, 6, 8
Immigrants serving the corrupt system	10 tokens	Voters, immigrants' vote, human caterpillars	Anger, contempt	7

Occurrences of *illegal immigrants* have also been identified in contexts where they are seen as a tool being used by the corrupt system for seeking immigrants' vote. The table above summarizes these thematic axes with contexts and/or lexical units found in our data and which help identify the main lexical fields of crime, unlawfulness, disaster, dirt and preferential treatment associated with immigrants:

Such contextual environments, lexical fields and collocates correspond to the entailments of the parasite metaphor which cause and reinforce negative emotions against the out-group, while enhancing ingroup cohesion and the ethos of power of the party. The ingroup cohesion is reinforced by Golden Dawn's persecuting hatred. This persecuting hatred is being developed by the construction of their enemies and the self-victimization.

In Greece the absence of any law to ban racist discourse contributes to the dehumanization of the out-group.

4.2 *From a Defensive Resentment to (Lost) Pride: The National Front*

The crisis experienced by France has also produced opportunities for 'charismatic leaders' who are ready to supply cheap messages of hope and re-stabilization (Eatwell 2007, 109; Williams 2011, 684).

However, in France, there is a law against hate speech, anti-semitic remarks and Holocaust denial even though the occurrences of Nazi references made by J.-M. Le Pen and his anti-immigrant stances are well-known (Davies 1999; Hainsworth 2004). Until the 1990's official party publications made overtly white supremacist and racist remarks (Davies 1999, 21), very similar to Golden Dawn's remarks. JMLP stated recently in 2005 that the Nazi occupation of France was "not particularly inhuman", a statement printed in a revisionist newspaper and for which he was put on trial; a few years earlier he had also pointed out that the holocaust had

been "a detail in history" (since he could not deny it legally). Indeed Castells's book (1998) demonstrates very convincingly the link between Le Pen and Hilter's tactics. These were cited as mere 'gaffes' on Le Pen's behalf (Williams *ibidem*, 684 and 690); as a form of apology, JMLP would consistently claim to say out loud what French people think and feel (Cambon 2006, 36),[24] his ideology being seen as masquerading as "everyday discourse and concerns" (Williams 2011, 683–684).

4.2.1 A Defensive Resentment

Indeed the National Front claims to always think of the national interest and therefore to preserve the interest of the people (Ceccaldi cited in Williams *ibidem*, 686) as the name of the party clearly states. Being the voice of those who do not dare or cannot speak is a tactic shared by other radical right parties around the world (Norris 2005). Our argument is that Le Pen, father and daughter, manipulate the imagery of past glory (Jeanne d'Arc, the tricolour of the French flag but the royal blue colour as well) and identity politics (what is to be 'pure' French) in order to focus the electorate on the emotion of Hubris (for JMLP) and Pride (for MLP).

Both have been digging into the 'origins' to assert the 'frenchness' of their party and give back some glory to a depressed electorate. Hence J.-M. Le Pen's tactic to take on as the party's 'muse' the emblematic Jeanne d'Arc, the famous 15th century French heroin, who fought against the foreign invaders. He has successfully promoted Jeanne d'Arc, 'the Saint of France', as an emotional rallying figure, which had been exploited before by different political actors to defend values dear to the extreme right such as patriotism, gallicism, anti-Semitism, imperialism and anti-feminism (Jennings 1994, 711). This choice is then useful especially at times of anti-immigration movement and potential national division. Jeanne d'Arc "has been proclaimed as the mother of French virtue and safeguard of national integrity" in the Front national rhetoric since the 1980s (Tumblety 2008, 13). The 'pucelle d'Orleans' was the focus of JMLP's speech, when conceding the party defeat after the first rounds of the 2008 and 2012 presidential elections. Defiant, Le Pen father had been staging the fête de la Pucelle for the last 40 years, on May 1, the day of the most traditional communist celebration.

This nationalistic stance also leads to the anxiety to define what is to be French, who is or who should be French, what are the other politicians doing for the French people (especially for the workers and the 'petite bourgeoisie'), as J.-M Le Pen's book title *Les Français d'abord* (French People First) testifies. JMLP Front national pursued for the last 40 years a defensive resentment against anything and anyone who seemed to stand in the way of France's autonomy and 'pure frenchood'.

[24] JMLP speaking of himself : Or de plus en plus nombreux sont les Françaises et les Français qui trouvent que Le Pen dit tout haut ce que tout le monde pense tout bas.

4.2.2 Dignity and Lost Pride

Although Marine Le Pen (MLP) maintains the important anti-immigration stance for an extreme right party (cf. Lubbers et al. study on extreme right voting), this has shifted into an anti-Muslim stance (but not an anti-semitic position), adopting a positioning which is different from her father's and by far dissimilar to Golden Dawn's, although indirectly the agenda is the same:

> Thus Muslims have taken the place held yesterday by the Jewish, the Arab or the migrant in the FN dialectics. Let us make no mistake: those who speak of Islamization are guided by the same xenophobic obsession that those who denounced the Judaization of our country in the 1930s.[25]

MLP would not, as JMLP did, play on words with the Nazi discourse (*solution finale* becoming *résolution finale*).[26] Her followers would less likely call leftists 'filth' as they did openly before (Tumblety 2008, 11). To detach herself even more symbolically from her father's stance, Marine Le Pen's presidential campaign was not carried out under the party's name but under the personalised label 'Rassemblement Bleu Marine' ('Marine Blue Rally'), competing directly with the conservative party (Sarkozy's UMP). Blue is not only the traditional colour of France with red but also a *royalist* colour, a choice which traditionally recalls the cult of origins specified in Kristeva's thesis. Her followers on forums are actually referring to her as Marine, playing on words with the colour bleu marine and her name.

She pursues a strategy which started in the 1990s when the Front national was contesting any relation with Nazism, facism or racism of all varieties (Williams 685). To do so, euphemisms have been invented to hide the party's agenda: we read 'national preference' instead of 'anti-immigrant', 'border control issues' or 'security' replace 'law and order' (Pégard 2002). All these euphemisms are changing the focus of the ideological positioning, sliding towards the Self, the nationhood, and sliding away from the Others, the immigrants, the foreigners, Them. This repositioning of the Self (especially in triggering the emotions of pride and dignity) helped the FN with Marine Le Pen to broaden the support base of the party in general, while gaining part of the traditionally communist electorate and female voters (William *ibidem*, 691, Shields 2013).

Pride is indeed the strong emotion promoted in the MLP's speeches.

As we have also noted in Sect. 2.1 in the Nazi discourse, Billig and Musolff noted that Pride is strategically very important in galvanizing the people into taking

[25] Ainsi le musulman a pris la place tenue hier par le juif, l'Arabe ou l'immigré dans la dialectique frontiste. Ne nous y trompons pas : ceux qui parlent de l'islamisation de la France sont guidés par la même obsession xénophobe que ceux qui dénonçaient la judaïsation de notre pays dans les années 1930.

[26] http://www.sbs.com.au/dateline/story/transcript/id/601463/n/Marine-Le-Pen

action, to avoid paralysis with too much fear.[27] The main action in MLP's speech is about *resistance* especially against the EU, the French public clearly rejecting a deeper integration; Le Pen (father) was describing his followers as "an *ardent and generous* minority" (Tumblety 2008, 10). Pride requires self-evaluation and self-representation (Tracy et al. 2010, 164), hence keeping Jeanne d'Arc as a symbol would benefit Marine Le Pen.

4.2.3 Speeches and Interviews

Marine Le Pen's speech and allure (the way she appears on the stage for instance) attest indeed to a new and dynamic style, in terms of appearance and discourse, building her image around lightness, blond hair and a smiling face, proud and perhaps more dignified than her father's. As interviewees explained why they attended one of her rallies:

(13) Because with her, it is different. She does not have her father's fascist ideas. She is much simpler.[28]

(14) She is much more prudent. She doesn't say that. She wants to be part of the next governing right.[29]

(15) There was no racist remarks to be overheard in the crowd (Williams 2011, 690).

Marine Le Pen is structuring *her speech*, appearance and actions as the dignified Saviour of the Nation:

> The national Front has spent the last few years regrouping, reconsidering its position, and gently manoeuvring a newer, younger, and, frankly, more beautiful face into the limelight. That new face belongs to Marine Le Pen (Wildman 2010).

She is less focusing her speech on the construction of the Other as evil, targeting above all on the trilogy of emotions Pride-Dignity-Hope, which will be the basis of acceptance of many of the arguments put forward in the Front national political platform. Marine Le Pen herself commented on her campaign as a rally of all French people:

(16) We have *imposed our themes* in this election and we have surprised everyone by the strength of our views, by their coherence. We *have become the centre of gravity* of French politics. We are the party of *reconciliation* for *all* French people. *Long live the nation! Long live the republic! Long live France!*(our italics).

(17) To *regain our sovereignty*. The people are the only *legitimate sovereigns*. Today, they have sold our sovereign liberty, little by little to the European Union, to the technocrats (our italics).

[27] We will not discuss pride in the discourse of Golden Dawn although it is also very much present, we will concentrate on how Marine Le Pen strategically uses this emotion.

[28] http://www.sbs.com.au/dateline/story/transcript/id/601463/n/Marine-Le-Pen

[29] http://www.sbs.com.au/dateline/story/transcript/id/601463/n/Marine-Le-Pen

In the three interviews we have studied (see references), statistically with the AntConc programme we have noticed that:

- for instance in her Henin Beaumont speech (March 2012) – the nouns (*France* and *French*) were more frequent than *immigration* (*France, French* are present 16 and 14 times respectively, *immigrants* is only heard twice and *immigration* once)*;*
- the word *fight* (résistance) was used with the collocates such as c*hildren (*enfants)*,* (*French*) *people* (les gens, les Français), the *nation, civilization* (civilization)*, mothers* (les mères), *values* and *truth* (la vérité): *children, people* were present seven times in her speech and *civilization, mothers, truth, value* (s) four times.

An excerpt of that speech in March 2012 exhibits the two cornerstones of a nationalistic stance as described by Kristeva and identified as well in Golden Dawn's speeches, namely:

- the return to a past glory:

> French civilization is a splendid alchemy of our art, our laws, our literature, of our fundamental rights, dearly acquired, of our beliefs, our values, our traditions, our habits, our mores, our code, our life-style [...].[30]

- and the emotion of pride:

> I think that the French should, at least, have the courage to vote for themselves, for their children (...) I say that *the spirit of France is not dead.* I assert that France can return to the path of *success and grandeur,* that *people can return to the path of pride, dignity and freedom.*[31]

In this sample of her speech Marine Le Pen presented herself as fighting more FOR values such as frenchness, family, nation, truth, than against a migrant or ethnic community (the word 'value' being preeminent in her speeches). This is also asserted in her followers' explanations who see "protection" and "freedom" when they are supporting Le Pen: "Because every citizen has a country and every citizen must defend his country".

Notably those values come before expected collocates such as *corruption* (the Front national criticizing the political decadence of the French Republic) and *immigration* which is still found most frequently with the qualifiers *massive* and *out of control* (incontrôlée), and therefore very negatively evaluated.

This new ideological distance is also true in the way she is portrayed in the French press.

[30] La civilisation française est une alchimie splendide de nos arts, de nos lois, de nos lettres, de nos droits fondamentaux chèrement acquis, de nos croyances, de nos valeurs, de nos traditions, de nos habitudes, de nos mœurs, de nos codes, de notre mode de vie [...].

[31] Je pense que les Français doivent enfin avoir le courage de voter pour eux-mêmes, pour leurs enfants, et ne plus voter pour tout donner à d'autres. [...] Je dis que l'esprit de la France n'est pas mort. J'affirme que La France peut retrouver le chemin de la réussite et de la grandeur, le peuple retrouver le chemin de la fierté, de la dignité retrouvée et de la liberté.

4.2.4 Written Data in the French Press

The statistics on the Marine Le Pen corpus show that the word 'extrême' (extremist) (extrême = 14 occurrences) is twice less frequent than 'right-wing' (droite = 29 occurrences) which could be interpreted as a successful 'recentrage' of the National Front:

Total number of word types: 1,899		Total number of word tokens: 10,506
Linguistic expression	Frequency	Number of occurrences
Droite	46	29
extrême	99	14

Other lexical findings show the shift of the FN focus when comparing with Sarkozy's results. We used the specificity criterium as well as the frequency parameter.

For the research corpus of Sarkozy concerning the two-month period before the vote (April-May 2012) the frequency and specifity of the lexical items in his data are the following: the word *immigration* is present 50 times and is found in the 26th place as far as frequency is concerned; moreover, expressions such as *viande halal* (halal meat), *immigration légale* (legal immigration) are high on the specifity score i.e. 21.01 for each and in the 50th place in terms of specificity. However, for Marine Le Pen, the word *immigration* is 14 times present in her data with a specificity of 11.25, half the score found in Sarkozy data. This shows how much the FN agenda was taken by Sarkozy campaign and how little difference the electorate would then see between what supposed to be a centre-right party (the Sarkozy UMP) and an extreme right party (the FN of Marine Le Pen). Indeed, and as mentioned before, unlike her father, she does not argue as much against immigration but against globalist elites. She also argues for national independence rather than national identity.[32]

Indeed, the most *specific* collocate found in MLP data with the name Marine Le Pen is dédiabolisation 'de-demonization':

Linguistic expression	Specificity
dédiabolisation 'de-demonization'	40.97

If we look further at other collocates which are the most specific to the corpus dedicated to Marine Le Pen, we will not observe any reference to extremism, immigration or law and order but to *démondialisation* 'deglobalization' and

[32] However, because of Sarkozy's push on the right during the electoral campaign, Marine Le Pen went back to a strong anti-immigration stance especially in November-December 2011 to keep her electorate happy and not to leave for the UMP, Sarkozy's party. The associations *diplômés étrangers* (foreign graduates) and *immigration légale* (legal immigration), appeared during that brief period as specific collocations in her data.

protectionnisme 'protectionism' i.e. to defend France against Europe and the world, as well as, indirectly, against Islam (with the noun *laïcité* 'secularism'):

Linguistic expression	Specificity
Démondialisation 'deglobalization'	63.18
protectionnisme (européen) '(European) protectionism'	54.02
laïcité 'secularism'	28.71

As far as emotions are concerned from March until December 2011 the word *angoissés* (worried, anguished) was among the most frequent adjectives in Marine Le Pen's data; however, in the April-May 2012 data, *orgueil* 'pride, excessive pride' is the only emotion to be ranked in her specificities (specificity 38.08) given that the FN was the third party in the first round of elections.

The articles relating to Marine Le Pen and her speeches highlight her positive qualities during her campaign. These collocations focus on her claims for *originality*, being an atypical party and representative and fighting with *courage* and *integrity* (systematic attack against the political corruption), the collocates 'corrupted ground', 'socialist mafia' and 'right-wing wheeler' being specific (with the same score) to this research corpus:

Linguistic expression	Specificity (highest score 78.02)
terre de corruption	23.86
mafia socialiste	23.86
droite affairiste	23.86

Like her father she asserts the existence of a conspiracy against her own party and against France (with Europe for instance and especially) and calls for *resistance* and *moral duty*. The imaginary is an identification with the Resistance and the image of Jeanne D'arc as already mentioned. This resistance will give the community the *sense of control* over their "destiny" against the immigrants, globalist elites, hooligans, etc. Indeed, among the most frequent collocates with the name *Marine Le Pen* are the adjectives *révolté* 'rebelled' and *républicaine* 'republican'.

Total No. of Collocate Types: 21		Total No. of Collocate Tokens: 95
Linguistic expression	Rank	Frequency
révolté	7	6.18473
républicaine	8	3.27783

Her claims for *popular legitimacy* are founded on the "gap" between the People and its representatives and the FN became the droite populaire 'labour right-wing' as well as a rallying party as the specific collocates during her campaign:

Linguistic expression	Specificity
droite populaire 'labour right-wing'	45.84
droite 'right –wing'	45.01
pôle de rassemblement 'rallying point'	28.47

Her motto is indeed to “gather” all the French people (pôle de rassemblement), encouraging them to take ‘the same’ pride in being French as these quotations show:

(18) Il faut rassembler les Français sans distinction d’origine (we have to gather all French people whatever their origin is)

(19) Il faut rassembler les Français et accéder au pouvoir (to gather the French people and to get the power)

(20) “(. . .) Qu’on soit homme ou femme, chrétien, juif, musulman ou non croyant, hétérosexuel ou homosexuel, on est d’abord Français!” a-t-elle lancé devant la statue de Jeanne d’Arc (man or woman, Christian, Jewish, Muslim, or non believer, heterosexual or homosexual, we are first French ! she claimed in front of the Jeanne d’Arc statue).[33]

The last quotation is even more remarkable that Le Pen father is openly against gay marriage, notwithstanding her embrace of the Muslim community which had been often the target of the antiracialistic ideology of the FN. The National Front president does not hesitate either to bury vindictiveness (hitherto implacable on behalf of JMLP) against the man who had sold off the Empire and French Algeria: “De Gaulle was able to make the people proud,” Marine Le Pen can assert without flinching. It has even been argued that Marine Le Pen uses arguments, and a stance reminiscent of this historic icon dear to the French electorate, De Gaulle himself (Onfrey 2013).

4.2.5 Neo-Racism?

However, it has been noted (Wildman 2010) that this new openness can be interpreted as neo-racism. As a matter of fact, her possibly feminist stance which Marine Le Pen adopts against the burqa/veil/hijab or against the Sharia law may also be a comfortable way to attack the Muslim population: “she often makes what might be called a feminist argument again for extreme-right wing politics” (Wildman *ibidem*). In the previous section we have also observed that references to secularism (i.e. a value of the French Republic) refer in an indirect and negative way to Islam and the several burqa/veil affairs in France.

As well, the FN leader speaks of immigration, she differentiates on criteria of *cultural* and *religion differentiation*, a strategy which looks more ‘legitimate’ to the electorate than physical or race discrimination. As a matter of fact, a concept which has been gaining popularity on the FN forum discussions is “racialisme” (racialism). This concept explains the reason why white separatist identity politics of Marine le Pen are argued not to be racist: separation from some communities is necessary because of cultural or social incompatibility. This separatism being based on social and cultural differences and not on ethnical hierarchy, can present itself as completely different from *racism*. Van Dijk also noticed that in the Spanish press

[33] http://frenchfries2012.blogspot.com/2011/05/marine-mise-sur-les-jeunes.html

the word *immigration* is especially associated linguistically with "alledged *cultural* threats" (2006c, 64).

Therefore, on the presumption of the existence of different "cultures" or "societies", we see a 'newspeak' of the extreme right, a "new racism" (Barker 1981) or "neoracism" (Balibar 1988) taking shape, much more subtle than Le Pen father's brutal stance. We could argue that in the case of Le Pen father, pride was blatantly 'hubristic', as we can observe for instance in his overuse of his own name (Cambon, 2006, 41) and in discursive strategies linked to the inferiorization of the Other. This 'hubristic pride' coupled with an overt racial discrimination (a person not being white-skinned cannot be a 'real' French person) makes him a blatant racist. However, with this 'neoracist' stance, Marine le Pen does not present the Roms or the migrants as biologically inferior but irretrievably different and therefore incompatible. These differences (the clothes, the beard but essentially the religion),[34] real or imaginary, appear as natural as the 'races': the 'pure race' has shifted into a 'genuine culture'(Taguieff 1988). This neoracism has become democratic and respectable since it is described even in left-wing newspapers as 'modernized' parties. Hubris has been made 'legitimate', and functions under the mask of pride, since it does not lead to hierarchy of civilization but to incompatibility of social groups.

5 Conclusion

Comparing both parties' strategies allowed us to bring out many similarities; they both share a similar nationalistic stance claiming to be a resistance movement; both of them call back to the origins reminiscent of Kristeva's analysis of hyper nationalism; both play on the semi-God(ness) positioning of the leader of each party, and they both use very similar means of persuasion (emotions). Both parties and both leaders have the same aim of ostracizing, functioning into legitimizing dominance (Memmi 1982), playing with the dichotomization (We-Them), the generalization, and the essentialization of the Other.

However, we have also identified the existence of two different strategies, namely defensive resentment (Pride) for the FN vs. persecuting hatred (Contempt) for Golden Dawn. Golden Dawn and the new FN using different emotions, achieve thereby a very different *ethos*.

Golden Dawn's pervasive discourse of exclusion calls for a persecuting hatred of anyone not being Greek or not adopting the party's ideology, by using and abusing the emotion of contempt against the Other. It is clear that in its racist stance, Golden

[34] See Marine Le Pen's recent outburst about the freed hostages commenting their 'scarfs', and 'the way they cut their beard' implicitly referring to Islam.

http://www.france24.com/en/20131031-france-hostages-strange-beards-clothes-trouble-marine-le-pen.

Dawn has chosen the inferiorization process. Its leader, Michaloliakos, surfing on the wave of despair and desperation caused by the dire economic crisis in Greece, openly asserts hatred towards all immigrants and all those who support them using metaphors such as parasites or rodents, well known to be very much part of the Nazi discourse. The ethos of Saviour is transformed into the ethos of Exterminator; Golden Dawn followers' habit of travelling in packs while wearing dark clothes, their logo resembling so much the Nazi swastika and Golden Dawn's secret army constitute the main symbolic ingredients of that ethos. In contrast, Marine Le Pen's strategy aimed, because of a different socio-historical context, to construct a defensive resentment by opting for the differentiation process. The National Front's agenda has indeed been reframed by MLP into a differentiation *modus operandi*, while constructing an ethos of Pride for the followers, an ethos of a new 'Frenchness' which could lead, although *indirectly*, to contempt of the Other. In her attempt to widen her audience and gain the women's vote (she succeeded in doing so in 2012, cf. Shields 2013), she focuses on France as a Nation (against Europe) while carefully using the anti-Muslim stance, distinguishing them from other communities (such as the Jewish community).

This study may also show that the more openly in conflict is the society, the more real becomes the target of bad feelings as we have recently seen in Greece: concepts cannot be abstracted from the hegemonic context within which they are embedded (Besnier 1990; Pavlenko 2005). Moreover, public discursive practices of 'state violence' seem to become more vested into the private affective meaning of words (cf. Kecskes private and public facets of lexical meaning), since in mainstream Greek journalistic discourse the word *immigrant* is often a collocate with *insecurity* for instance and thereby perpetuating, maintaining or installing social fears and anxiety, emotions well-known to govern subjects (cf. Deleuze and Foucault works). The on-going crisis sees a resurgence of such rhetoric in many countries (open affiliation to Nazism, economic or political refugees becoming the cause of the financial crisis, etc.). Politicians will play even more on the urgent need for ethnic/ national survival. And if we believe MacMillan's recent thesis in *The War That Ended Peace* published last year, this is extremely worrying given that right-wing and nationalist sentiments had been a factor before the First World War.

References

Amossy, R. (2008). Modalités argumentatives et registres discursifs: le cas du polémique. In L. Gaudin-Bordes & G. Salvan (Eds.), *Les registres. Enjeux pragmatiques et visées stylistiques* (pp. 93–108). Louvain-la-Neuve: Academia-Bruylant.

Angouri, J. (2010). 'If we know about culture it will be easier to work with one another': Developing skills for handling corporate meetings with multinational participation. *Language and Intercultural Communication, 10*(3), 206–224.

Averill, J. R. (1980). A constructivist view of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research and experience* (pp. 305–339). New York: Academic.

Baider, F. (2012). Le sentiment 'haine' en contexte linguistique et cognitif. *ÉtudesRomanes de Brno, 33*(2), 171–192.

Baider, F. (2013). Hate: Saliency features in cross-cultural semantics. In I. Kecskes & J. Romero-Trillo (Eds.), *Linguistic aspects of intercultural pragmatics* (pp. 7–27). Berlin: Walter de Gruyter.

Baider, F., & Jacquey, E. (2014). Data mining for tracking gender bias: The 2012 socialist primaries duo. In H. Tyne et al. (Eds.), *Ecological and data-driven perspectives in French language studies* (pp. 222–245). Newcastle upon Tyne: Cambridge Scholar Press.

Balibar, E. (1988). Y a-t-il un "néo-racisme"? In E. Balibar & I. Wallerstein (Eds.), *Race, Nation, classe, Les identités ambiguës* (Vol. 1, pp. 27–41). Paris: La Découverte.

Banks, D. (2007). Ideology, context and text in a systemic functional model. In G. Girard (Ed.), *Texte(s), contexte(s), hors-texte(s)* (pp. 179–191). St Etienne: Presses Universitaires de St Etienne.

Barcelona, A. (2000). Introduction. The cognitive theory of metaphor and metonymy. In A. Barcelona (Ed.), *Metaphor and metonymy at the crossroads* (pp. 1–28). Berlin/New York: Mouton de Gruyter.

Barker, M. (1981). *The new racism*. London: Junction Books.

Bednarek, M. (2008). *Emotion talk across corpora*. New York: Palgrave Macmillan.

Besnier, N. (1990). Language and affect. *Annual Review of Anthropology, 19*, 419–451.

Billig, M. (1978). *Fascists: A social psychological view of the national front*. London: Harcourt Brace Jovanovich.

Blumenthal, P. (2002). Profil combinatoire des noms. Synonymie distinctive et analyse contrastive. *Zeitschriftfür Französische Sprache und Literatur, 112*, 115–138.

Burke, K. (1984). The rhetoric of Hitler's 'battle'. In M. Shapiro (Ed.), *Language and politics* (pp. 61–80). Oxford: Blackwell. [Originally in *The Southern Review*, *5*, (1939/40): 1–21.]

Cambon, E. (2006). Le Pen disant "Le Pen" ou la construction du chef charismatique par lui-même. In D. Banks (Ed.), *Aspects linguistiques du texte de la propagande* (pp. 31–42). Paris: L'Harmattan.

Castells, R. (1998). *Hitler, Le Pen, Maigret, leur programme*. Paris: Ed. Raymond Castells.

Charteris-Black, J. (2004). *Corpus approaches to critical metaphor analysis*. Basingstoke: Palgrave-Macmillan.

Charteris-Black, J. (2005). *Politicians and rhetoric. The persuasive power of metaphor*. Basingstoke: Palgrave-Macmillan.

Chilton, P. (2005). Manipulation, memes and metaphors: The case of Mein Kampf. In L. de Saussure & P. Schulz (Eds.), *Manipulation and ideologies in the twentieth century* (pp. 5–45). Amsterdam/Philadelphia: John Benjamins.

Conger, J. A. (1991). Inspiring others: The language of leadership. *Academy of Management Executive, 5*(1), 31–45.

Cornell, D. (1995). What is ethical feminism. In S. Benhabib, J. Butler, D. Cornell, & N. Fraser (Eds.), *Feminist contentions: A philosophical exchange*. London: Routledge.

Davies, P. (1999). *The national front in France: Ideology, discourse and power*. New York: Routledge.

Deleuze, G., & Foucault, M. (1973). *The intellectuals and power: A discussion between Gilles Deleuze and Michel Foucault*. New York: Telos Press.

Drouin, P. (2003). Term extraction using non-technical corpora as a point of leverage. *Terminology, 9*(1), 99–117.

Eatwell, R. (2007). Towards a new model of right-wing charismatic leadership/Hacia un nuevo modelo de liderazgo carismático de derecha. In M. A. Simon Gomez, J. Antón Mellón, & P. Perrineau (Eds.), *La extrema derecha en Europa desde 1945 a nuestros dias* (pp. 19–38). Madrid: Tecnos.

Frijda, N. H. (2007). What emotions might be? Comments on the comments. *Social science information sur les sciences sociales, 46*, 433–443. doi:10.1177/05390184070460030112.

Frye, M. (1983). *The politics of reality: Essays in feminist theory*. Freedom: Crossing Press.

Groenendyk, E. (2011). Current emotion research in political science: How emotions help democracy overcome its collective action problem. *Emotion Review, 3*(4), 455–463.
Hainsworth, P. (2004). The extreme right in France: The rise and rise of Jean-Marie Le Pen's Front National. *Representation, 40*, 101–114.
Hart, D., & Matsuba, M. K. (2007). The development of pride and moral life. In J. L. Tracy, R. W. Robins, & J. P. Tangney (Eds.), *The self-conscious emotions: Theory and research* (pp. 114–133). New York: Guilford Press.
Hawkins, B. (2001). Ideology, metaphor and iconographic reference. In R. Dirven, R. Frank, & C. Ilie (Eds.), *Language and ideology. Vol. II: Descriptive cognitive approaches* (pp. 27–50). Amsterdam/Philadelphia: John Benjamins.
Holmes, M. (2004). Feeling beyond rules: Politicizing the sociology of emotion and anger in feminist politics. *European Journal of Social Theory, 7*, 209–227.
Jennings, E. (1994). Reinventing Jeanne: The iconology of Joan of Arc in Vichy schoolbooks, 1940–44. *Journal of Contemporary History, 29*, 711–734.
Kahan, D., & Nussbaum, M. (1996). Two conceptions of emotion in criminal law. *Columbia Law Review, 2*, 270–374.
Kecskes, I. (2008). Dueling context: A dynamic model of meaning. *Journal of Pragmatics, 40*(3), 385–406.
Kecskes, I. (2011). Interculturality and intercultural pragmatics. In J. Jackson (Ed.), *The Routledge Handbook of Intercultural Communication* (pp. 67–84). London: Routledge.
Kecskes, I. (2012). Encyclopaedic knowledge and cultural models. In H.-J. Schmid (Ed.), *Cognitive pragmatics* (Handbook of Pragmatics, Vol. 4, pp. 175–200). Berlin/New York: de Gruyter, Mouton.
Kemper, T. D. (2001). A structural approach to social movement emotions. In J. Goodwin, J. M. Jasper, & F. Polleta (Eds.), *Passionate Politics* (pp. 58–73). Chicago: University Chicago Press.
Kitayama, S., & Masuda, T. (1995). Reappraising cognitive appraisal from a cultural perspective. *Psychological Inquiry, 6*, 217–233.
Kohl, D. (2011). The presentation of "self" and "other" in Nazi propaganda. *Psychology & Society, 4*(1), 7–26.
Koselak, A. (2005). Mépris/dédain, deux mots pour un même sentiment? *Lidil, 32*, 21–34.
Kövecses, Z. (1986). *Metaphors of anger, pride and love: A lexical approach to the structure of concepts* (Pragmatics & beyond VII: 8). Amsterdam/Philadelphia: John Benjamins.
Kristeva, J. (1993). *Nations without nationalism*. New York: Columbia University Press.
Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought* (2nd ed., pp. 202–251). Cambridge: Cambridge University Press.
Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: The University of Chicago Press.
Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh*. New York: Basic Books.
Lazarus, R. (1991). *Emotion and adaptation*. New York: Oxford University Press.
Lazarus, R. (1997). Acculturation isn't everything. *Applied Psychology, 46*, 39–43.
Lesage, R., Wilson, P., Bissonnette, C., & Drouin, P. (1993). Enquête sur l'état d'utilisation des outils automatisés d'aide à la rédaction dans les organisations. *META, 38*(2), 367–389.
Lubbers, M., Gijsberts, M., & Scheeper, P. (2002). Extreme right-wing voting in Western Europe European. *Journal of Political Research, 41*, 345–378.
Lyons, W. (1985). *Emotion*. Cambridge: Cambridge University Press.
MacMillan, M. (2013). *The war that ended peace*. United-Kingdom: Random House LLC.
Marcus, G. E. (2000). Emotions in politics. *Annual Review of Political Sciences, 3*, 221–250. doi:10.1146/annurev.polisci.3.1.221.
Marlin, R. (1998). *The David Levine affair: Separatist betrayal or McCarthyism North?* Halifax: Fernwood.
Memmi, A. (1994 [1982]). *Le racisme*. Paris: Gallimard.
Mio, J. S. (1997). Metaphor and politics. *Metaphor and Symbol, 12*, 113–133.

Mio, J. S., Riggio, R. E., Levin, S., & Reese, R. (2005). Presidential leadership and charisma: The effects of metaphor. *The Leadership Quarterly, 16*, 287–294.

Musolff, A. (2007). Which role do metaphors play in racial prejudice? The function of anti-Semitic imagery in Hitler's 'Mein Kampf'. *Patterns of Prejudice, 41*(1), 21–44.

Musolff, A. (2008). What can critical metaphor analysis add to the understanding of racist ideology? Recent studies of Hitler's anti-semitic metaphors. *Critical Approaches to Discourse Analysis Across Disciplines*, *2*(2), 1–10. [online]. http://cadaad.org/ejournal

Norris, P. (2005). *Radical right: Voters and parties in the electoral market*. New York: Cambridge University Press.

Onfrey, M. (2013, December 10). La fausse mue du vrai serpent. *Le Monde*. http://www.lemonde.fr/idees/article/2013/12/10/le-pen-la-fausse-mue-du-vrai-serpent_3528583_3232.html

Papaioannou, K. (2013). Τα "καθαράχέρια" της ΧρυσήςΑυγής. Εφαρμογές ναζιστικής καθαρότητας. [The clean hands of Golden Dawn. Applications of Nazi cleanliness]. Athens: Μεταίχμιο.

Pavlenko, A. (2005). *Emotion and bilingualism*. Cambridge: Cambridge University Press.

Pégard, C. (2002, April 25). Etat de choc: la faute à qui? *Le Point,* 8–13.

Perry, B. (2006). A crime by any other name: The Semantics of "Hate". *Journal of Hate Studies, 4*(1), 120–137.

Plantin, C. (2011). *Les bonnes raisons des émotions. Principes et méthode pour l'étude du discours émotionné*. Berne: Peter Lang.

Potvin, M., Morelli, A., & Mettewie, L. (2004). Racisme dans les Rapports entre Groupes Nationaux au Canada et en Belgique? *Canadian Ethnic Studies Journal, 36*(3), 25–60.

Rosebury, B. (2003). On punishing emotions. *Ratio Juris, 16*(1), 37–55.

Savage, R. (2007). Disease incarnate: Biopolitical discourse and genocidal dehumanisation in the Age of modernity. *Journal of Historical Sociology, 20*, 404–440.

Scherer, K. R. (2000). Emotions as episodes of subsystem synchronization driven by nonlinear appraisal processes. In M. D. Lewis & I. Granic (Eds.), *Emotion, development, and self-organization: Dynamic systems approaches to emotional development* (pp. 70–99). New York: Cambridge University Press.

Shamir, B., House, R. J., & Arthur, M. B. (1993). The motivational effects of charismatic leadership: A self-concept theory. *Organization Science, 4*, 1–17.

Shamir, B., Zakay, E., Breinin, E., & Popper, M. (1998). Correlates of charismatic leader behavior in military units: Subordinates' attitudes, unit characteristics, and superiors' appraisals of leader performance. *Academy of Management Journal, 41*, 387–409.

Shields, J. (2013). A change of style or of substance? *Parliamentary Affairs, 66*(1), 179–196.

Sinclair, J. M. (2004). *Trust the text: Language corpus and discourse*. London: Routledge.

Smith, C. A., & Lazarus, R. (1990). Emotion and adaptation. In L. A. Pervin (Ed.), *Handbook of personality: Theory & research* (pp. 609–637). New York: Guilford Press.

Taguieff, P. A. (1988). *La force du préjugé: essai sur le racisme et ses doubles*. Paris: La Découverte.

Tracy, J. L., Shariff, A. F., & Cheung, J. T. (2010). A naturalist's view of pride. *Emotion Review, 2*(2), 163–177.

Tumblety, J. (2008). Contested histories: Jeanne d'Arc and the front national. *The European Legacy: Toward New Paradigms, 4*(1), 8–25.

Tutin, A., Novakova, I., Grossman, F., & Cavalla, C. (2006). Esquisse de typologie des noms d'affect à partir de leurs propriétés combinatoires. *Langue française, 150*, 32–49.

Van Dijk, T. A. (1995). Ideological discourse analysis. In E. Ventola & A. Solin (Eds.), *New courant* (Special issue interdisciplinary approaches to discourse analysis, Vol. 4, pp. 135–161). Helsinki: University of Helsinki. http://www.discourses.org/OldArticles/Ideological%20discourse%20analysis.pdf.

Van Dijk, T. A. (2006a). Discourse, context and cognition. *Discourse Studies, 8*(1), 159–177.

Van Dijk, T. A. (2006b). Discourse and manipulation. *Discourse & Society, 17*(2), 359–383.

Van Dijk, T. A. (2006c). Racism and the press in Spain. Available at http://www.discursos.org/unpublished%20articles/Racism%20and%20the%20press%20in%20Spain.htm

Vaughan, E., & Clancy, B. (2013). Small corpora and pragmatics. In J. Romero-Trillo (Ed.), *Yearbook of corpus linguistics and pragmatics* 2013. New Domains and Methodologies (pp. 53–73). New York/Heidelberg: Springer.

Wierzbicka, A. (1988). L'amour, la colère, la joie, l'ennui. La sémantique des émotions dans une perspective transculturelle. *Langages, 89*, 97–107.

Wierzbicka, A. (1992). Defining emotion concepts. *Cognitive Science, 16*(4), 539–581.

Wildman, S. (2010). Neo-nazi no longer? Marine Le Pen tones down France's far right. http://www.politicsdaily.com/2010/05/31/neo-nazis-no-longer-marine-le-pen-tones-down-frances-far-right/

Williams, M. H. (2011). A new era for French far right politics? *Análise Social, XLVI*(201), 679–695.

Willner, A. R. (1984). *The spellbinders: Charismatic political leadership*. New Haven: Yale University Press.

Windisch, U. (1978). *Xénophobie? Logique de la pensée populaire*. Lausanne: L'âge d'Homme.

Corpus

Greek oral data:

First speech: 6th May 2012 after their electoral victory of the first round in parliamentary elections. (http://www.youtube.com/watch?v=E4AXJx3IzdY)

Second speech: 17th June 2012 after the electoral triumph of Golden Dawn of the second round.

Third speech: 16th June 2013 celebrating the continuing rise of Golden Dawn. http://www.xryshaygh.com/index.php/enimerosi/view/h-omilia-tou-archhgou-ths-chrushs-aughs-sthn-giorth-gia-ths-eklogikhs-nikhs#.Ub72vOemi8E

Speech by E. Zaroulia: 18th October 2012. http://www.youtube.com/watch?v=6Y09xr846cg

Speech by Kasidiaris, the spokeperson for Golden Dawn. http://www.makeleio.gr/index.php?option=com_content&id=4025:--------video&Itemid=2

French corpus:

Racism and Jean-Marie Le Pen: JMLP and the Arabic community. http://www.youtube.com/watch?v=q3pApd53-hk

Jean-Marie Le Pen and the Roms. http://www.lemonde.fr/societe/article/2013/12/19/jean-marie-le-pen-condamne-a-5-000-euros-d-amende-pour-des-propos-sur-les-roms_4337485_3224.html. Jean-Marie Le Pen sur l'immigration en France – 27-01-1988. http://www.youtube.com/watch?v=tRT8GruN__o

Jean-Marie Le Pen sur l'immigration en France - 27-01-1988 http://www.youtube.com/watch?v=tRT8GruN__o

La haine de le PEN contre les Algériens. http://www.youtube.com/watch?v=1JxSpwh4gPU

Marine Le Pen speeches during the 2012 campaign and Interviews to the press: http://www.sbs.com.au/dateline/story/transcript/id/601463/n/Marine-Le-Pen

Marine Le Pen, FN à Nice le 11/09/2011. http://www.youtube.com/watch?v=YlqTuiGjLo0

Point de vue de Jean-Marie Le Pen sur l'immigration en France. http://www.youtube.com/watch?v=WYuVxGzJh84

Débat sur l'identité nationale: Marine Le Pen commente. http://www.youtube.com/watch?v=zH094DHN0KY

Part III
Advances in L2 Corpus-Based Pragmatics Research

‘*We Went to the Restroom or Something*’. General Extenders *and Stuff* in the Speech of Dutch Learners of English

Lieven Buysse

Abstract This chapter investigates how learners of English who are native speakers of Dutch use general extenders such as *and stuff* and *or something*. The corpus consists of the Dutch component of the Louvain International Database of Spoken English Interlanguage (LINDSEI), which is comprised of 50 interviews of some 15 min each. These data are compared with the Louvain Corpus of Native English Conversation (LOCNEC), LINDSEI’s native speaker reference corpus. The study shows that overall frequencies of general extenders point at a close alignment of the two speaker groups, but that discrepancies exist if these numbers are further broken down for the adjunctive and disjunctive categories of general extenders. The former type is used considerably less frequently in the learner corpus than in the native, whereas the opposite holds for the latter. A detailed qualitative and quantitative analysis offers a few tentative explanations for the learners’ choice of general extenders, most notably L1 transfer, the intensity of exposure to certain forms in the target language, and learners’ restricted repertoire of pragmatic devices.

Keywords General extenders • Vague language • English as a foreign language • Learner English • Learner corpora • Foreign language learning

L. Buysse (✉)
Faculty of Arts, KU Leuven – University of Leuven, Campus Brussels, Warmoesberg 26, 1000 Brussels, Belgium
e-mail: lieven.buysse@kuleuven.be

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2, DOI 10.1007/978-3-319-06007-1_10, © Springer International Publishing Switzerland 2014

1 Introduction

General extenders[1] have been defined as utterance-final tags that “have nonspecific reference or ‘general’ reference, and [. . .] ‘extend’ otherwise grammatically complete utterances” (Overstreet and Yule 1997: 251). In English they typically take the form of a conjunction (*and/or*) plus a vague noun phrase (Overstreet and Yule 1997: 250; Cheshire 2007: 156), such as *and things (like that), and all that stuff* and *or something*. Because they fulfil (interpersonal) functions in spoken language similar to those of prototypical members of the class of pragmatic markers, they have often been considered a subset within this class, with a clearly distinguishable structure (see e.g. Dubois 1992: 181; Brinton 1996: 32; Aijmer 2002: 211; Overstreet 2005: 1846; Terraschke 2007: 142).[2]

A common distinction, first pointed out by Overstreet (1999), in analyses of general extenders is that between adjunctive forms, which start with *and*, and disjunctive forms, which start with *or*. In accordance with the nature of these two coordinators, the two types of general extenders that they preface each have a distinct functional scope. Adjunctive general extenders indicate that “there is more” (Overstreet 2005: 1851), and therefore serve as a cue for the hearer to infer further instantiations of the same category as that which immediately precedes the general extender. As such, it is a token of intersubjectivity and asserts common ground between speaker and hearer: the speaker assumes that the hearer is able to make this inference on the basis of shared knowledge or experience. Disjunctive general extenders, on the other hand, point at the existence of possible alternatives to the referent of the prior phrase or clause. These are, therefore, often used to hedge violations of Grice’s maxim of quality (Overstreet 1999: 112) by indicating that the prior co-text should not necessarily be taken as an accurate proposition but, for example, as an approximation of what is meant.

Since the 1980s general extenders have received a great deal of attention in the analysis of native English discourse (e.g. Dines 1980; Channell 1994; Overstreet 1999; Aijmer 2002, 2013; Cheshire 2007; Tagliamonte and Denis 2010). This expanding body of research has not only fine-tuned the distinction between adjunctive and disjunctive forms, but has also complemented the initial emphasis on ideational functions with one on interpersonal functions (Overstreet 1999) and has provided evidence of both ongoing grammaticalization and changing preferences of use (Cheshire 2007; Tagliamonte and Denis 2010). As has been observed

[1] These items are also known under a plethora of competing terms, such as “set-marking tags” (e.g. Dines 1980) and “vague category identifiers” (e.g. Channell 1994), an overview of which can be found in Tagliamonte and Denis (2010: 335–336). I have chosen to adopt the term coined by Overstreet and Yule (1997), *general extenders*, because it has gained by far the widest currency in the field and does not (over)emphasize any particular function that such items can fulfil.

[2] There is no shortage of alternative terms for what are here referred to as *pragmatic markers* either. As a consequence, there is considerable variation in how these authors label the umbrella category containing such items as *you know, I mean, so, well,* the question tag *innit*, *okay, look*, and general extenders. It is clear from their descriptions of this category, though, that what they refer to is roughly the same.

for pragmatic markers in general (Müller 2005; Buysse 2012), and even for most pragmalinguistic devices (Callies 2013) for that matter, the use of general extenders by language learners has largely remained in the shade. This study aims to address this dearth by analysing data of EFL learners who have Dutch as their mother tongue, and by comparing them with native speaker practice.

2 The Learner Perspective

The language research community always understandably observes some lag time between the first analyses of a certain phenomenon in the standard variety of the language and the expansion of the domain to include other language varieties, such as learner language. A few studies have included general extenders in – mostly quantitative – overviews of pragmatic markers or vague expressions, such as Hasselgren (2002) for Norwegian secondary-school pupils, Aijmer (2004) for Swedish learners of English and De Cock (2004) for French-speaking EFL learners. Others have provided more elaborate accounts albeit so far for a highly restricted number of L1s: Terraschke (2007, 2010) and Terraschke and Holmes (2007) have worked with data from learners with German L1, and Parvaresh et al. (2012) with native speakers of Persian.

Most studies on general extenders in learner English have attested considerably lower frequencies for these forms in the learner data than in the native (Hasselgren 2002; Aijmer 2004; De Cock 2004; Fernandez and Yuldashev 2011; Parvaresh et al. 2012). Terraschke's (2007, 2010) analyses of the discourse of native speakers of German forms a notable exception to this otherwise seemingly robust tendency, in that she finds that these learners overall make more frequent use of general extenders in general but their repertoire spans a smaller number of forms. Some forms appear more frequently in Terraschke's learner corpus because learners project functions fulfilled by cognate forms in German onto English forms.

De Cock (2004) notes a similar phenomenon. Contrary to all the other general extenders that she lists, *and so on* is more prevalent in the learners' speech than in the native speakers' although this form is more typical of formal types of speech. This, De Cock (2004) claims, "only adds to the impression of detachment and formality [learners] may well give in informal situations" (2004: 236). As with most pragmatic devices, all too (in)frequent use or even misuse of general extenders is unlikely to result in misunderstandings or a communication breakdown, but the impression learners give to their interlocutors may indeed be a less favourable one.

Given the fledgling status of learner corpus research on general extenders, it would be untimely to draw hard and fast conclusions about how learners use these forms on the basis of this sketchy literature survey, especially because the settings from which corpora have drawn tend to differ dramatically, as do L1 backgrounds and native speaker groups with whom comparisons are made. For example, Terraschke (2007) reports a higher prevalence of disjunctive than adjunctive forms in her corpus of conversations of native speakers of German, whereas Fernandez and Yuldashev (2011) observe the opposite in their corpus of instant messaging conversations among EFL learners of varying L1 backgrounds. Also the yardstick against which learner

practice is measured should be taken into account in comparing results from different studies, because as Aijmer (2013) has shown, different native speaker varieties of English display different preferences in their use of general extenders: Terraschke (2007) uses native speakers of English from New Zealand, De Cock (2004) compares with native speakers from the United Kingdom, Fernandez and Yuldashev (2011) from the United States, and Parvaresh et al. (2012) do not use any yardstick.

3 Data and Methodology

To investigate the use of general extenders in the speech of EFL learners who are Belgian native speakers of Dutch the Dutch component of the Louvain International Database of Spoken English Interlanguage (LINDSEI; Gilquin et al. 2010) was selected, consisting of 50 interviews with learners of English who were majoring in English in their second or third year at university. These participants can hence be considered as having arrived at the final stage of their formal education in English, and a spot check of overall proficiency levels of learners in this component of LINDSEI has shown that they can be situated at an advanced level (Gilquin et al. 2010: 10–11). The results for this learner corpus will be juxtaposed with those for the Louvain Corpus of Native English Conversation (LOCNEC), which is the native speaker reference corpus of LINDSEI and consists of 50 interviews with undergraduates from a British university. Table 1 presents the key figures of the two corpora.

Both corpora were composed along the same format. Each interview lasts about 15 min, and starts with a part in which the interviewee is expected to talk independently about a topic (such as a book, a film, a travel experience) for about 2 min, which leads to a conversation with the interviewer. The interviews end with a brief picture-based story-telling part.

The general extenders were extracted from the corpora with the concordance application of Wordsmith Tools 5.0 (Scott 2008) on the basis of forms that surfaced in the prior investigations mentioned in Sects. 1 and 2: vague nouns such as *things, stuff, something* were searched for, automatically sorted and the general extenders were subsequently manually detected.[3] A list of the resulting forms is provided in Table 2 (see Sect. 4). All examples have been taken from the corpora under investigation; the transcription conventions can be found in the Appendix. In all

Table 1 Size of the Dutch learner corpus (DU) and the native corpus (NS)

	DU	NS
Interviews	50	50
Total words	100,454	170,533
Total words interviewees	79,652	125,666
av. words interviewees	1,593	2,513

[3] A similar approach has been adopted by Aijmer (2013: 132).

quantitative data mentioned in this study both absolute and relative numbers will be provided. The relative numbers represent the number of tokens per 10,000 words of interviewee speech. In comparing the two corpora apparent differences between relative frequencies have been tested for statistical significance with independent-samples t-tests in the statistical software programme SPSS. This method takes into account that the corpora consist of clearly distinguishable independent texts from different language users (viz. 50 interview(ee)s each), and can therefore account for inter-learner variability in the learner corpus.

4 A Bird's Eye View of General Extenders in the Learner and Native Corpus

In both the learner and the native corpus almost 26 tokens of general extenders occur per 10,000 words (Table 2). This remarkable parallel between the two corpora does not hold, though, if we zoom in on the categories of general extenders. The adjunctive forms comprise 61 % of all general extenders in the native corpus, the disjunctive 37 %, and those forms which belong to neither because they are not prefaced by a coordinator take up a marginal 2 %. Exactly the same percentages are noted for the Dutch learner corpus, but here the disjunctive forms pick up the majority share. This results in highly significant differences between the corpora (indicated with asterisks in Table 2) for both the adjunctive (NS: M = 16.35, SD = 15.17; DU: M = 9.03, SD = 12.45; $t(98) = 2.64$, $p = 0.010$)[4] and the disjunctive (NS: M = 9.82, SD = 8.19; DU: M = 15.34, SD = 14.60; $t(77.08) = -2.33, p = 0.022$) categories as a whole; no noteworthy differences can be reported for the residual category (NS: M = 0.37, SD = 1.32; DU: M = 0.54, SD = 2.04; $t(98) = -0.52, p = 0.603$) and for the overall numbers of general extenders (NS: M = 26.54, SD = 19.87; DU: M = 24.92, SD = 20.15; $t(98) = 0.40, p = 0.687$).

For convenience sake those general extenders that are variants of the same form and fulfil identical functions (such as *and stuff* and *and stuff like that*) have been considered together in this overview. In Sect. 5 the adjunctive forms that yield statistically significant differences between the learners and native speakers will be discussed in greater detail, and Sect. 6 will do the same for the disjunctive category.

5 Adjunctive General Extenders

As shown in Table 2, the adjunctive general extenders as a group occur with a significantly higher frequency in the native corpus than in the learner corpus. Some forms only occur in the learner corpus, and could be classified as "non-native-like":

[4] The reported statistical output data are: the mean (M) and standard deviation (SD) for each corpus, the *t*-value with degrees of freedom in brackets, and the *p*-value indicating statistical significance if it is lower than 0.05.

Table 2 Absolute (n) and relative (per 10,000 words) frequencies of general extenders in the Dutch learner corpus and the native speaker corpus

	DU		NS	
	n	rel.	n	rel.
and so	2	0.25	0	0.00
and so on	20	2.51	2	0.16
and (all) things (like that)***	9	1.13	84	6.68
and stuff (like that)	28	3.52	49	3.90
and (all) that	3	0.38	7	0.56
and (all) that kind/sort of thing	1	0.13	9	0.72
and all that (kind/sort of) stuff	3	0.38	3	0.24
and all those kind of stuff	1	0.13	0	0.00
and everything (like that)**	7	0.88	44	3.50
and something like that	0	0.00	1	0.08
and whatever	2	0.25	0	0.00
Total adjunctive**	**76**	**9.54**	**199**	**15.84**
or anything (like that)	11	1.38	25	1.99
or something (like that)**	91	11.42	72	5.73
or so***	16	2.01	4	0.32
or stuff like that	1	0.13	0	0.00
or that kind of stuff	2	0.25	0	0.00
or whatever*	4	0.50	21	1.67
Total disjunctive**	**125**	**15.69**	**122**	**9.71**
(all) stuff like that	2	0.25	2	0.16
(all) that kind of stuff	2	0.25	0	0.00
all that	0	0.00	1	0.08
whatever	1	0.13	2	0.16
Total other	**5**	**0.63**	**5**	**0.40**
Overall total	**206**	**25.86**	**326**	**25.94**

Significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

and so is probably motivated by its formal resemblance to both the Dutch general extender *en zo* (which functionally corresponds with *and things* and *and stuff*) and the English form *and so on*; the two tokens of *and whatever* are likely to be due to confusion with *or whatever*; and the one instance of *and all those kind of stuff* reflects a grammatical problem. These forms do not occur with any meaningful frequency, though, and have been included in the calculations of the prevalence of adjunctive general extenders because (i) they do fulfil the function of an adjunctive general extender, and (ii) their influence on quantitative results is negligible. Two adjunctive general extenders yield statistically significant differences between the two corpora: the forms with *and things* (NS: M = 7.26, SD = 9.43; DU: M = 1.25, SD = 4.25; $t(68.119) = 4.107$; $p = 0.000$) and those with *and everything* (NS: M = 3.45, SD = 4.91; DU: M = 0.87, SD = 2.57; $t(73.954) = 3.29$, $p = 0.002$). These will, therefore, be discussed in the following sections. If a chi-square test is performed instead of a t-test (thereby abstracting from the fact that each corpus consists of 50 separate interviews), the difference between NS and DU for *and so on* also reaches the threshold of statistical significance, and will

hence also be addressed below.[5] Although no statistically significant differences hold for the forms with *and stuff*, these will also be included in the discussion of the forms with *and things* given their strong resemblance.

5.1 And So On

The odd one out in this category is beyond any doubt *and so on*, which occurs significantly more often in the learner data than in the native (2.15 tokens per 10,000 words versus 0.16, respectively), thereby defying the dominant trend in this category. In native speaker discourse *and so on* is typical of written language (Biber et al. 1999: 117) and formal speech (Overstreet 1999: 7), neither of which applies to the setting of the present corpus, which qualifies as that of an "informal interview" (Gilquin et al. 2010: 8). This general extender is, therefore, virtually absent from the native speaker corpus, but appears 20 times (2.51 times per 10,000 words) in the learner corpus, which seems to corroborate De Cock's (2004) findings for the French component of LINDSEI (see Sect. 2). Two important differences ought to be mentioned, though. First, the frequencies reported by De Cock (2004: 237) and Buysse (forthcoming) for this form in the English speech of native speakers of French are three times as high as those in the Dutch learner corpus. Second, the distribution rates are clearly different, in that only 6 Dutch-speaking interviewees make use of *and so on*, as opposed to 23 French-speaking interviewees (Buysse forthcoming). This signifies that the numbers for the Dutch learner corpus have been heavily influenced by two interviewees who use *and so on* as a pet form (or in Gilquin's (2008: 129) terms: as a "pragmatic teddy bear"), one of whom is responsible for nine tokens and another for seven tokens. This explains the discrepancy in the results between the two statistical tests: the differences between the corpora are only considered significant if the corpus is regarded as one long stretch of speech (chi-square) and not if the 50 individual interviews for each corpus are taken into account (t-test).

Excerpt (1) contains two instances of *and so on*. The second of these illustrates its prototypical function of a list completer, whereas the first does not complete a list but merely follows a single example to signal that other instantiations of the same category should be taken into account as well. The latter function is much more common in the Dutch learner corpus, although it is more typically taken on by other adjunctive general extenders such as *and stuff* or *and things*.

[5] To perform this test the two-sample corpus frequency test wizard was used (http://sigil.collocations.de/wizard.html), which determines automatically whether chi-square or log-likelihood is the more appropriate test (Hoffmann et al. 2008: 85). The difference was found to be significant at $p < 0.001$.

(1) <Iee> the journalists . would . (er) after th= seeing . th= the press screening and so on would (er) . apply for (erm) . well would come to me and ask for interviews . well via email **and so on** . </Iee>
<Ir> <overlap/> oh yeah </Ir>
<Iee> <overlap/> and I had to (eh) <overlap/> arrange it </Iee>
<Ir> <overlap/> interesting </Ir>
<Iee> with the distributors . with the m = film distributor distributors and the managers **and so on** </Iee>
(DU09)

5.2 And Stuff (Like That) versus And (All) Things (Like That)

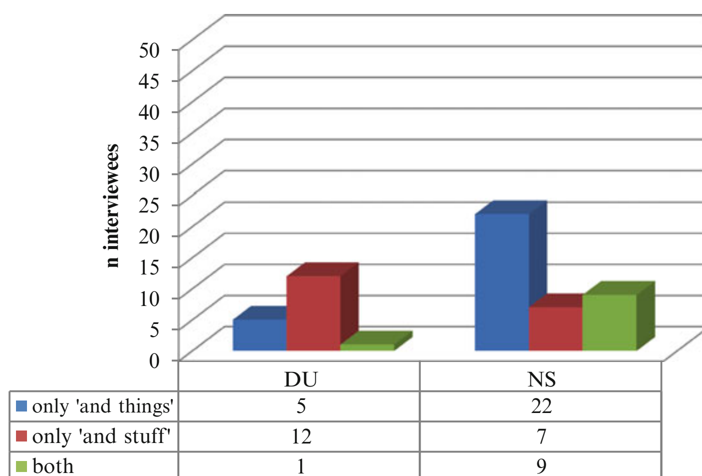
And stuff and *and things*, and their more elaborate forms, are probably the most frequent adjunctive general extenders, and therefore most often play their typical double role to indicate that "there is more" (Overstreet 2005: 1851) as well as to build rapport with the interlocutor by establishing common ground. As Terraschke and Holmes (2007) point out, the referential and attitudinal functions occur virtually simultaneously, making it impossible to separate the two. In (2), for example, the interviewee mentions a component of the English Language course module and indicates with *and stuff* that this is just one component. At the same time it subtly voices the assumption that the interviewer should be capable of coming up with further examples as she too is highly familiar with the curricula at that university.

(2) <Iee> and . why I chose . linguistics cos it sounded more .. language orientated </Iee>
<Ir> uhu </Ir>
<Iee> and English language <overlap/> is like history **and stuff** yeah </Iee>
(NS03)

Similarly, the learner interviewee in (3) suggests with *we stayed in hotels and stuff* that the interviewer should be able to conjecture that the former spent a while in one place, staying in a hotel and engaging in other activities that could be done there, as anyone with some knowledge of travel arrangements would know.

(3) <Iee> [. . .] and then we stayed a week with them and then we . toured a little bit . we went with (er) the car . and then we stayed in hotels **and stuff** . and then we went a week to . Disney World </Iee>
(DU40)

An interesting contrast between the two corpora is found in the frequencies for the forms containing *and things* and *and stuff*. These take up first and second position in the ranking of adjunctive general extenders in the native corpus, and language users appear to exhibit a modest preference for either (Fig. 1): 31 interviewees use a form



	DU	NS
only 'and things'	5	22
only 'and stuff'	12	7
both	1	9

Fig. 1 Distribution range of *and things (like that)* and *and stuff (like that)* in the Dutch learner corpus and the native corpus

containing *and things* and 16 one containing *and stuff*, but only 9 of these use both and even then with an outspoken dominance of one of these.

In the Dutch learner corpus the relative frequency of *and stuff (like that)* approaches that of the native corpus (3.52 and 3.90, respectively), but the relative frequency of *and (all) things (like that)* is significantly lower (1.13 and 6.68, respectively). Like in the native corpus, interviewees exhibit a distinct preference for either form (Fig. 1): 13 interviewees prefer *and stuff (like that)* and 6 prefer *and things (like that)*, one of whom uses both (once each).

Contrary to what has been reported for British English (e.g. Aijmer 2002: 221; Aijmer 2013: 134), *and things (like that)* has not been found with any particularly high frequency in American English (e.g. Overstreet 2005: 1848; Aijmer 2013: 134). The propensity for *and stuff (like that)* in the learner corpus, as compared to the relatively restricted use of *and (all) things (like that),* may be due to a more extensive exposure of learners in the Dutch-speaking part of Belgium to American English, where forms with *and things* have not been reported to be highly common. Since the supply of television series and films in the media immediately available to the learners is dominated by the United States and these are not dubbed but subtitled, they are more likely to have become familiar with *and stuff*-forms than with *and things*-forms. For the last few decades *and stuff (like that)* has, however, been on the rise in the speech of (adolescent) Brits too (Cheshire 2007; Palacios Martínez 2011), which may explain the fairly high frequency of both forms in the (British) native corpus, with a preference for *and things*.

Since *and things* and *and stuff* have identical functions and most language users opt for one or the other, a glance at the combined frequencies for these forms may be revealing as to the extent to which groups of language users employ pragmatic devices with these ideational and interpersonal functions. If these frequencies are

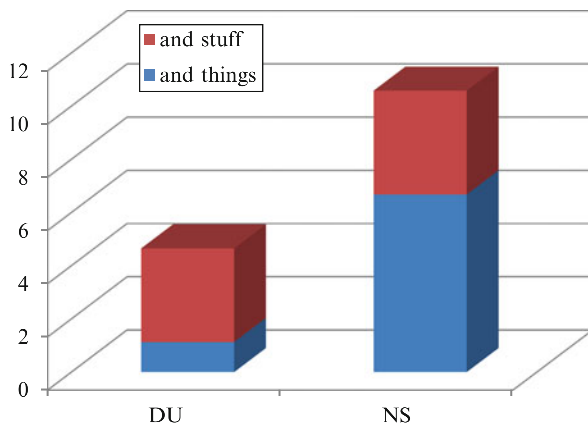


Fig. 2 Combined relative frequencies (per 10,000 words) for *and things (like that)* and *and stuff (like that)* in the Dutch learner corpus and the native corpus

indeed added up for the native corpus, the dominance of these forms becomes even more apparent, with 10.58 tokens per 10,000 words out of an overall relative frequency of 15.84 adjunctive general extenders. The perceived gap with the learners, which was somewhat masked by the frequency for *and stuff (like that)*, widens to a chasm (Fig. 2): with only 4.65 tokens per 10,000 words, the forms *and stuff/things (like that)* remain well below half the frequencies in the native corpus. It should be noted, though, that the learners have *and so on* perform functions typical of *and stuff/things* (see Sect. 5.1), which may have a minor influence on these numbers as well.

A final noteworthy discrepancy between the two corpora for these general extenders is that the native speakers display an outspoken preference for the short forms *and things* and *and stuff* with ratios around 60 % (57 % and 65 %, respectively) compared to the longer variants. In the Dutch corpus the balance tips the other way, with an overall ratio of 38 % for short forms (only 2 out of 9 take the form *and things*, and 12 out of 28 take the form *and stuff*). Bearing in mind Aijmer's (2002) suggestion that extensive use of short forms can be regarded as a sign of automaticity in a language user's repertoire, this observation could be taken as an indication that the language learners who make use of these general extenders have not yet automatised this use to the same extent as their native speaker peers. Another factor may be that learners use more words as a strategy to gain reflection time for what to say next or how to best express what they would like to say next.

5.3 *And Everything*

Like *and stuff* and *and things*, *and everything* is primarily a general extender with a set-marking function combined with the interpersonal function of building rapport with the interlocutor. Its meaning is, however, not compositional, in that it does not incite the hearer to think of all conceivable members of the inferred set, contrary to what one might gather from the semantic meaning of *everything* (Ward and Birner 1993).

This is the third most prevalent adjunctive general extender in the native speaker corpus, yet it only surfaces sporadically in the learner corpus, viz. 7 times – or 0.88 times per 10,000 words – spread across 6 interviews. The set-marking function of *and everything* is exemplified in (4):

(4) <Ir> . okay (em) do you have any hobbies outside of school that you like anything you spend a lot of time . doing </Ir>
<Iee> (erm) it's <overlap/> really </Iee>
<Ir> <overlap /> in your free time </Ir>
<Iee> busy for (er) now now . nowadays with school **and everything** so I really don't have a lot of hobbies [. . .] </Iee>
(DU02)

Additionally, *and everything* may have an intensifying function (Overstreet 1999; Overstreet and Yule 2002), when it marks "an extreme value on whatever salient scale had been evoked" (Overstreet and Yule 2002: 788). As such it serves to evaluate an element as "remarkable, surprising, or (a maximum) extreme" (Overstreet 1999: 146). This function does not figure prominently in the present corpora, neither in the learner nor in the native component. One of the rare instances that come close to it can be found in excerpt (5), where the interviewee reminisces about the romantic scenery of a lake. *And everything* could here act as an invitation to the interviewer to imagine even more idyllic images.

(5) <Iee> that was very beautiful that was really impressive in terms in terms of the beauty .. and when I went it was a black beautiful summer's evening <XX> <overlap/> all the light reflecting on the lake </Iee>
<Ir> <overlap/> oh yeah </Ir>
<Iee> **and everything** </Iee>
(NS13)

Three tokens in the learner corpus take the form of the formulaic sequence *X and everything, but Y*, as identified by Overstreet and Yule (2002). This formula allows speakers to anticipate certain expectations among their hearers, and to "offer a justification for thinking contrary to those expectations" (2002: 792). For example, in (6) the interviewee describes the venue of a language summer course as "a very nice place", including a swimming pool and a great view, but then she adjusts the exclusively positive image by adding as a downside that it is very expensive.

(6) <Iee> it is a <overlap/> very nice place </Iee>
<Ir> <overlap/> with swimming pools and (er) </Ir>
<Iee> with a swi = swimming < overlap /> pool and </Iee>
<Ir> <overlap/> <XX> </Ir>
<Iee> . great view **and . everything** but </Iee>
<Ir> (mhm) </Ir>
<Ir> it's very expensive to go there and the food is very expensive the[i:] </B>
(DU17)

6 Disjunctive General Extenders

As a group the disjunctive general extenders focussed on in this study, are significantly more frequent in the learner corpus than in the native, with 15.69 tokens per 10,000 words in the former and only 9.71 in the latter. Two disjunctive forms are learner-specific in this corpus (Table 2), viz. *or stuff like that* and *or that kind of stuff*, but these were only used once and twice, respectively. The former has been attested in other native speaker corpora (see e.g. Evison et al. 2007: 143; Aijmer 2013: 137), whereas the latter has not and its utterance may have been motivated through analogy with the adjunctive form *and that kind of stuff*. As indicated for the adjunctive forms, these learner-specific disjunctive general extenders have been included in the calculations of the prevalence of disjunctive forms because (i) they fulfil the function of a disjunctive general extender, and (ii) their influence on quantitative results is negligible. For three disjunctive general extenders the threshold for statistical significance was reached (see Table 2), viz. for the forms with *or something* (NS: M = 5.88, SD = 6.06; DU: M = 11.16, SD = 12.41; $t(71.11) = -2.70$; $p = 0.009$), *or so* (NS: M = 0.29, SD = 1.02; DU: M = 2.19, SD = 5.52; $t(52.30) = -2.40$; $p = 0.020$), and *or whatever* (NS: M = 1.63, SD = 3.43; DU: M = 0.50, SD = 1.83; $t(74.85) = 2.05$; $p = 0.044$). These will be discussed in the following sections.

6.1 Or Something (Like That)

The only (disjunctive) general extender with a truly meaningful frequency in the learner corpus is *or something (like that)*. It is almost all by itself responsible for the significantly higher prevalence of the disjunctive category, and takes up 44 % of all general extender tokens in the learner corpus (amounting to 11.42 tokens per 10,000 words, compared to 5.73 in the native corpus). Its high prevalence in the corpus as well as its versatility (as reported by *inter alia* Overstreet 1999) warrants a closer scrutiny to this general extender's functions than to any other's. Inspired by Overstreet's (1999) discussion of the main functions of *or something (like that)*, the following sections will, therefore, be devoted to the functions of set-marking, speculation, various forms of approximation (recollections, reported speech, amounts or numbers, and word choice), analogies, and jokes. Attention will also be paid to cases in which interviewees confuse *or something (like that)* with *or anything (like that)*, and in Sect. 6.1.10 absolute and relative frequencies will be provided for each of the attested functions and for both groups of interviewees (see Table 3).

6.1.1 Set-Marking

The prototypical function of *or something* is to "encourage the hearer to think of an alternative to some members of the set" (Aijmer 2002: 218). In some cases this can be taken quite literally, viz. when the speaker indicates that the prior element is merely an illustration of a more general point. For example, in excerpt (7) going to a beach stands for the set of 'activities to spend a lazy holiday', and could therefore be replaced by related activities such as lying at the swimming pool. In a similar vein, the profession of doctor in (8) represents the set of 'venerable professions any parent would want their child to practise', and could have been replaced by, for example, that of lawyer or chemist.

---

(7) <Iee> and I would prefer to go somewhere or go to a place and see what there is to see about that place </Iee>
<Ir> (mhm) </Ir>
<Iee> rather than just going and just laying<?> on a beach **or something** and </Iee>
<Ir> and then go out at night and drink </Ir>
(NS07)

(8) <Iee> so parents still want their child to to study </Iee>
<Ir> (mhm) </Ir>
<Iee> and become a doctor or **or something** </Iee>
<Ir> yeah yeah yeah </Ir>
<Iee> instead of . w= while (em) if you're a good . (er) craftsman </Iee>
(DU14)

In this function *or something* sticks closest to its compositional, ideational meaning in which the meaning of *or* combined with that of *something* is sufficiently transparent to determine the meaning of the structure as a whole. It is also the function that most closely resembles the primary adjunctive general extenders (e.g. *and things*, *and stuff*): both indicate that the preceding element is just an instantiation of a more general category, but while adjunctive forms suggest that a longer list of examples could be forged, disjunctive forms suggest that the element(s) provided could be replaced by another element. These forms are, therefore, not interchangeable. It would, for example, not be possible to use *and stuff* instead of *or something* in (7) and (8), because you cannot lie on a beach and at a swimming pool at the same time nor is it likely for a student to become a doctor and a lawyer at the same time.

6.1.2 Speculation

As Overstreet (1999: 112) contends, *or something* (and disjunctive general extenders in general) can serve as a hedge on the Gricean maxim of quality, in that it marks the preceding utterance as potentially inaccurate. This may take the shape of speculation, which is illustrated in excerpts (9) and (10). In the former the interviewee hypothesises that if a car got clamped although it had a parking permit,

it could have been an expired permit, and in the latter the interviewee voices her fear of what might have happened to her little brother if there had not been a parasol beneath the window to break his fall. In both cases the speculative nature of the utterances is not only marked by the general extender but also by modal auxiliaries expressing uncertainty or hypothesis.

(9) <Iee> where you can like pay and display (erm) .. (eh) and if people don't do that and they haven't got a permit at all then they will get clamped yes </Iee>
<Ir> cos I saw that one of the cars had .. a permit on its </Ir>
<Iee> yeah </Iee>
<Ir> windscreen so </Ir>
<Iee> I don't know it it could have been from a different year **or something** or it could have been just <overlap/> <XXX> </Iee>
(NS18)

(10) <Iee> [. . .] if he would have slided down . the other way of the <overlap/> parasol </Iee>
<Ir> <overlap /> (mhm) </Ir>
<Iee> he would (er) would have crushed a wall **or something** and then then he . would <laughs> be no more (em) . either but I think he he has a a very . strong guardian angel so (er) </Iee>
(DU28)

6.1.3 Approximate Recollections

Overstreet (1999) indicates that, still as a hedge on the quality maxim, a disjunctive general extender and *or something* in particular, "functions to mark an utterance, or part of an utterance, not just as potentially inaccurate but as an approximation" (1999: 115). The utterance referred to may take different forms, the first of which is when a speaker is unable to convey the preceding element accurately because they do not fully remember it. In (11), for example, the interviewee recounts an anecdote of seeing Al Pacino in the flesh on Broadway but is uncertain as to why the actor was there, whereas in (12) the interviewee attempts to remember exactly what was on a T-shirt.

(11) <Iee> yeah <laughs> there was this group of people outside a theatre apparently he was doing a stage play **or something** </Iee>
<Ir> mm </Ir>
<Iee> and so oh what are those people looking at we went over and it was Al Pacino . I was amazed <X> took photos of Al <laughs> it was great </Iee>
(NS53)

(12) <Iee> but (em) . I did . once I . in September I went to: London <overlap /> with some friends of mine . and (em) . we were actually just wa= . wandering down we we we got to a sort of (er) a square . and all people together there and (em) they were wearing these tee-shirts (em) . national youth theatre **or something** </Iee>
(DU22)

Overstreet (1999) does not explicitly distinguish approximate recollections as a sub-category of *or something*, but it clearly surfaces in the present corpus and is sufficiently distinct from the other functions to warrant its inclusion in this discussion. It is different from set-marking tags because no category from which alternatives could be plucked seems easily conceivable. Recollections can be set off from speculations as well, in that the speaker has actually experienced what s/he is trying to remember, so the utterance that *or something* marks scores higher on a scale of certainty when recollections are concerned than with speculations.

6.1.4 Approximate Reported Speech

When a speaker wants to convey someone else's words, this rarely takes the form of a verbatim report of what that person actually said. To give a "more dramatic version of the event" (Overstreet 1999: 119) stretches of reported speech or thought can, therefore, be marked by the speaker with *or something* with the meaning 'or something to that effect'. The interviewee in (13), for example, recounts the story of a film he has seen and at one point explicitly marks a segment as a quotation with *or something*.

(13) <Iee> so then . he goes back to . he ge= and he gets beaten up by the police and he goes back to the house . of (er) . well he to him it's just a house where he can go for help but it is <?> actually <?> one of the people he attacked so <X> he says oh come in come in realises who it is locks him in the locks him in the basement <X> locks him in the[i:] attic rather and plays Beethoven really loud so of course that's cracking <?> him <?> up so he jumps out of the window tries to kill himself and then .. he doesn't kill himself he breaks almost every bone in his body it all turns political <u>and the government say</u> oh it was a terrible thing to do blablablablabla er here's a <?> big stereo system **or something** <overlap/> and (er) </Iee>
(NS54)

It is clear that the interviewee in this example does not quote anyone literally, because he claims to be quoting a vague institution ("the government") and he signals leaving out a passage ("blablablablabla"). The addition of *or something* only intensifies the message that the hearer should not take the preceding stretch of reported speech literally.

Also in (14) the interviewee explicitly marks two upcoming quotations with a reporting verb (*to say*) and ends both with *or something* to signal that they should not be taken at face value. Note that in this excerpt the quotations cannot be regarded as verbatim renderings of what someone has actually said either, since they are situated in a hypothetical context.

(14) <Iee> I I think so because for instance when . (er) . (er) . parents say . (er) when a child comes up to his parents <u>and he says</u> I: want to go . to a technical school **or something** </Iee>
<Ir> (mhm) </Ir>
<Iee> <u>the parents say</u> are you sure shouldn't you: </Iee>
<Ir> (mhm) </Ir>

(continued)

<Iee> study: or or </Iee>
<Ir> (mhm) </Ir>
<Iee> shouldn't you: . try </Iee>
<Ir> (mhm) </Ir>
<Iee> (er) Latin or </Iee>
<Ir> yeah sure (mm) </Ir>
<Iee> modern languages **or something** </Iee>
(DU14)

This function borders on approximate recollections insofar as speakers are ordinarily assumed to rely on their memory to include reported speech in their discourse. The tokens that have been included in this category, however, mark stretches of speech that need not have been uttered, as shown above. Moreover, they often serve a double function. As has been noted for other pragmatic markers that signal reported speech (see e.g. Jucker 1993 and Müller 2005 on *well*), it is not altogether clear whether *or something* should be considered as belonging to the reported speech segment or should be regarded as a tag added by the speaker to the reported speech. Indeed, in most cases these two options are plausible, as in the examples above where the tokens of *or something* could be regarded as set-marking tags with reference to the prior element. This is hardly surprising, because the reported speech has been fabricated by the speaker, and hence each element contained in it could be regarded as an approximation of what has really been uttered in the context referred to.

6.1.5 Approximate Amounts or Numbers

Probably the clearest type of approximation context in which *or something* occurs, is with amounts or numbers, in which case the general extender "marks the named amount [. . .] as an estimate, which may not be exactly right" (Overstreet 1999: 115). This function has been reported by *inter alia* Overstreet (1999), Aijmer (2002) and Koester (2007), and can apply to various elements, the most prominent of which are price, age, period, time, and a number of people. The former two are illustrated in excerpts (15) and (16).

(15) <Iee> <begin_laughter> I know it is <end_laughter> <X> and like sometimes it's it's only it's only when I'm at university that I go <X> the pictures a lot cos at home you think that's a fiver I can't afford that but up here you think <overlap/> well it's only a pound and a half </Iee>
<Ir> <overlap/> it's okay </Ir>
<Iee> **or something** </Iee>
(NS28)

(16) <Iee> yeah it was (er) .. (eh) rather recently I think and it was also a young author . who was only thirty years old **or something** </Iee>
(DU06)

6.1.6 Approximate Word Choice

When a speaker is not certain of the accuracy of a term, they may mark this with *or something* as being "close to correct" (Overstreet 1999: 116). In (17), for example, the interviewee seems to forge a term to denote a document and indicates that this is most likely not the designated term. In a similar vein, the interviewee in (18) probably means some sort of break room or relaxation room but uses the term *restroom* instead, which clearly does not convey the intended meaning. With *or something* he indicates awareness of the potential inaccuracy of this term.

(17) <Ir> isn't it automatic when you're ill and you say well <overlap/> could you give me one </Ir>
<Iee> <overlap/> I I don't really don't really know I mean one of my friends she was ill and then she had to <X> to the doctor's and fill out like a self-certificate **or something** and sign something to say that the department can actually have access to your medical records </Iee>
(NS37)

(18) <Iee> no it it was relaxing mostly in the evenings when <overlap/> we . </Iee>
<Ir> <overlap/> (mhm) </Ir>
<Iee> went to the rest (er) restroom **or something** and had fun with friends but (er) . in the day itself it was (er) yeah we had to hurry to get to that museum before it closed </Iee>
(DU45)

The hearer often gets additional clues of the speaker's apprehension in using a potentially inaccurate term, such as the pragmatic marker *like* in (17), the truncation and filled pause in (18), or the overly explicit clause indicating word-searching in (19), where the interviewee refers to playing the recorder but calls it the flute instead.

(19) <Iee> yeah we had to . the the: . I don't know what it's called . the flute **or something** . in in high school [. . .] </Iee>
(DU07)

6.1.7 Analogies

In attempting to describe a novel experience, a speaker may compare it to a concept that the hearer can be assumed to be familiar with, which Overstreet (1999: 119) dubs an analogy. Since the interviewer has not been to Australia, she does not know what Melbourne is like in excerpt (20). The interviewee expects her to be more familiar with London, which is what Melbourne is then compared to.

(20) <Iee> it's nicer than Melbourne cos I went there as well </Iee>
<Ir> oh yeah </Ir>
<Iee> and Melbourne's just a bit .. well like (er) London **or something** really </Iee>
(NS26)

The experience in (21) is not one from the past but is rather one that the interviewee envisages for herself in the future: she wants to become an on-screen journalist and to complete the profile that she has in mind she gives an example of a rolemodel, viz. the well-known BBC sports presenter Gary Lineker.

(21) <Ir> would you like to be . in front of the camera or behind the camera </Ir>
<Iee> in <overlap/> front is fine </Iee>
<Ir> <overlap/> in front . okay </Ir>
<Iee> like . the female . Gary Lineker **or** <**overlap/**> **something** </Iee>
(DU05)

Notice that in both excerpts the analogy itself is also prefaced by the pragmatic marker *like*, and that *or something* additionally emphasises that the analogy is not foolproof.

6.1.8 Jokes

Or something (like that) can play a role in jokes or cartoons, according to Overstreet (1999), where they serve as a "purposeful exaggeration [. . .] which should not be taken too literally" (1999: 120). The only two instances of this function in the present corpus are found in the same fragment from one interview with a native speaker. The interviewee claims not to have a talent for languages, and jokingly aims to make the cause for this apprehensible with a graphic description involving a personification of the interviewee's brain and head.

(22) <Iee> <X> I'm just not a language person my head is not geared towards speaking languages I think my science side of my brain's like taken <X> my head <X> just pushed the language behind my ear **or something** <**overlap/**> **like that** </Iee>
<Ir> <overlap/> <laughs> well I'm going to try and find <overlap/> it </Ir>
<Iee> <overlap/> yeah I'll have to I'll have <XX> searching on the floor and see if I can stuff it back in **or** <**overlap/**> **something** </Iee>
<Ir> <overlap/> we're going to dig <laughs> <X> to try to find it </Ir>
(NS31)

The interviewer's uptake, with laughter and an accommodating response to both instances, shows that the joking nature of these remarks has successfully been conveyed to the hearer.

6.1.9 Confusion with *Or Anything (Like That)*

The disjunctive general extenders *or something (like that)* and *or anything (like that)* occur "in complementary distribution between assertive and non-assertive contexts" (Channell 1994: 132), respectively. This reflects the grammatical distinction between the basic forms *some* and *any*, from which these general extenders have been derived (Biber et al. 1999: 176–177). In learner English the boundaries

between these two forms is not always as strictly adhered to as in native English. Take excerpt (23): the context is clearly non-assertive (indicated by "never"), yet the interviewee does not use *or anything* but the assertive form *or something*.

(23) <Iee> [. . .] and we went there and (em) we never really . ate in in (er) </Iee>
<Ir> <overlap/> (mhm) </Ir>
<Iee> in a restaurant **or something** just </Iee>
<Ir> (mhm) </Ir>
<Iee> because we're also .. (em) walking around the entire day and (em) . just quickly buying a sandwich or quickly buying a </Iee>
(DU49)

In the present corpus *or something* replaces *or anything* in 16 instances, which occur in 11 interviews. Dutch, the learners' mother tongue, does not have a distinction similar to that between *some* and *any* in English, which most likely caused these mistakes. Such instances have already been reported by De Cock (2004) for the French component of LINDSEI. She describes this practice as "idiosyncratic misuse of a target language sequence" (2004: 237), although doubt could be shed on whether this is truly idiosyncratic when it has been attested in different learner corpora (De Cock 2004; Buysse forthcoming), and for each of these in the speech not of a single speaker but of quite a few speakers.

6.1.10 Functions of *Or Something (Like That)* in Numbers

The distribution rates for *or something (like that)* are almost identical in the two corpora, with 34 of the 50 native speaker participants in this study using it at least once, and 33 of the 50 language learners. Contrary to our findings for the adjunctive general extenders *and (all) things/stuff (like that)*, learners and native speakers alike display an outspoken preference for the short form *or something*, with ratios of 82 and 76 %, respectively.

Table 3 shows that in both the learner and the native corpus the number one function of *or something (like that)* is approximation. There is, however, a considerable gap between the relative frequencies of this cluster of functions for the two corpora: it is 2.5 times larger in the learner than in the native corpus (NS: M = 2.52, SD = 3.43; DU: M = 6.59, SD = 9.55; $t(61.44) = -2.83$; $p = 0.006$). If the category "confusion with *or anything*" is left out of the equation because it does not as such represent a function of *or something (like that)*, approximations amount to over 70 % of all tokens of this general extender in the learner corpus. Within this cluster, a highly significant difference also holds for approximate word choice (NS: M = 0.27, SD = 1.09; DU: M = 2.30, SD = 4.65; $t(54.42) = -3.01$; $p = 0.004$). If a chi-square test is performed instead of a t-test (thereby abstracting from the fact that each corpus consists of 50 separate interviews, as was done for *and so on* in Sect. 5.1), the difference between NS and DU for approximate amounts or numbers reaches the threshold of statistical significance as well (at $p < 0.05$). Learners and native speakers have *or something* fulfil its set-marking function to a similar extent.

Table 3 Quantitative overview of the functions of *or something (like that)* in the Dutch learner corpus and the native corpus

	DU			NS		
	n	rel.	%	n	rel.	%
Set-marking	14	1.76	15.4	26	2.07	36.1
Speculation	7	0.88	7.7	9	0.72	12.5
Approximation**	53	6.65	58.2	33	2.63	45.8
Recollections	*4*	*0.50*	*4.4*	*8*	*0.64*	11.1
Reported speech	*9*	*1.13*	*9.9*	*8*	*0.64*	11.1
Amounts, numbers	*21*	*2.64*	*23.1*	*14*	*1.11*	19.4
*Form of lexical items***	*19*	*2.39*	*20.9*	*3*	*0.24*	4.2
Analogies	1	0.13	1.1	2	0.16	2.8
Jokes	0	0.00	0.0	2	0.16	2.8
Confusion with *or anything***	16	2.01	17.6	0	0.00	0.0
Total *or something* **	**91**	**11.42**	**100.0**	**72**	**5.73**	**100.0**

Statistical significance: $^{*}p < 0.05$, $^{**}p < 0.01$

It is remarkable, though, that its share in the learner corpus does not come anywhere near that in the native, which is due to (i) the apparent overweight of the approximation functions and (ii) the additional category “confusion with *or anything*” in the learner corpus.

How can the statistically significant differences between the corpora for *or something (like that)* be explained? First of all, none of these has been influenced by an exceptionally high frequency of these functions in a single interview. All functions were found in at least 12 learner interviews each.

Secondly, it stands to reason that confusion with *or anything* yields a highly significant difference, since this is a learner-specific use that is unlikely to appear with any meaningful frequency in a native speaker corpus.

Thirdly, the significantly higher prevalence among the learners of approximate amounts or numbers may have been caused by a limited inventory of pragmatic devices which the learners appeal to. Whereas native speakers often have a wide array of forms at their disposal to mark approximate numbers (such as the pragmatic marker *like*), learners stick to one option. The same is likely to hold for approximate word choice: pragmatic markers such as *like, sort of, kind of, you know* and *I mean* tend to be grossly underrepresented in learner English, whereas native speakers use them to the extent that they have been denounced by language purists as “verbal viruses” (see e.g. Berkley 2002). For example, in (24) *like* twice marks an upcoming phrase as approximate, as does *sort of* in (25).

(24) <Iee> I think it’s it’s a case that when I qualify I have to do a year to: **like** <u>validate</u> my degree you have to do **like** <u>a year’s probational teaching</u> </Iee>
(NS22)

(25) <Ir> can you can you actually visit it </Ir>
<Iee> (erm) we just went we just went inside the[i:] **sort of** <u>chapelly bit</u> (erm) it wasn’t a guided tour so yeah you can go inside </Iee>
(NS45)

Likewise, the disjunctive general extender *or whatever*, which can be used to much the same effect as *or something*, is barely used in the learner corpus as opposed to the native (see Sect. 6.3). This tentative explanation tallies with reports on the (marginal) use of pragmatic markers with similar functions in learner English in general (e.g. Hasselgren 2002; Romero-Trillo 2002; Aijmer 2004; Müller 2005; Mukherjee 2009; Buysse 2011).

Finally, an additional factor is indubitably at play for approximate word choice, viz. language learners are likely to face more problems of formulation than native speakers (see e.g. Aijmer 2013:145) and – certainly when they are advanced learners – they hence feel a greater need to mark lexical items as potentially inaccurate, which *or something (like that)* caters for.

6.2 Or So

Although this is the second most common disjunctive general extender in the Dutch learner corpus, it can hardly be regarded as particularly prevalent. It fulfils much the same function as *or something* as a numerical approximator (Channell 1994; Aijmer 2002; Koester 2007), and is hence in many cases interchangeable with it, as is exemplified in (26).

(26) <Iee> no I didn't .. but (em) . I heard from someone that (em) they have cut the story they've because it . lasted for . three hours **or so** </Iee>
(DU16)

Channell (1994: 59) aptly points out, however, that *or so* can only approximate numbers whereas the scope of *or something* is considerably wider, as we have seen in Sects. 6.1.3, 6.1.4, 6.1.5, and 6.1.6. Excerpt (27) shows that the language learners do not always adhere to this functional restriction, which is the case in 6 out of the 16 tokens of *or so* that have been identified in the learner corpus.

(27) <Iee> [. . .] so they're better in English but . when . like . there's a phone call **or so** .. first I have to say . like . hello it's <first name of interviewee> and then they say <X> . oh hi . how are you . in Dutch </Iee>
(DU40)

This may be attributed to a lack of awareness of this rule in the target language, which may have been intensified further by L1 transfer because *of zo*, the Dutch cognate form of *or so*, can be used for any type of approximation.

In one case *or so* even substitutes for *or anything* (see Sect. 6.1.9):

(28) <Iee> especially also in the way of living they .. it seems like they don't have any stress **or so** .. I think the climate and the sun also . helps with that because </Iee>
(DU33)

If these non-standard uses of *or so* are discounted in the comparison with the native corpus, the difference no longer achieves statistical significance (with a relative frequency of 1.25 in the learner corpus). The statistically significant difference for this form can, therefore, be attributed to non-native-like uses that confuse it with *or something/anything*.

6.3 Or Whatever

Or whatever can be employed with the same functions as *or something*, but additionally implies that "it doesn't actually matter whether certain details are exactly correct" (Overstreet 1999: 123). Although it ranks third in the list of disjunctive general extenders in the native corpus, it is only used by four learners. These four instances are either set-marking or approximate the form of a lexical item, as in excerpts (29) and (30), respectively.

---

(29) <Iee> (em) . that's what I do at home . usually . (em) if I could be <X> to go outside I just . call some friends and . go to . movies **or whatever** </Iee>
(DU25)

(30) <Iee> <overlap/> yes because (er) . all the . in the city . (em) the city centre is really . empty at night . </Iee>
<Ir> (mhm) </Ir>
<Iee> so all the cr = all the .. gangsters **or whatever** . </Iee>
<Ir> (mhm) </Ir>
<Iee> they all come to the centre and . they just rob . everyone who walks on the street so </Iee>
(DU12)

---

7 Conclusion

At a cursory glance, a quantitative juxtaposition of a Dutch learner corpus and a comparable native speaker corpus would lead us to believe that the learners border on native speaker practice when it comes to the use of general extenders in English. This apparent alignment of the overall frequencies masks an underlying chasm, though: the learners only scantily make use of adjunctive general extenders, and turn significantly more frequently to disjunctive general extenders than their native peers. Within these broad categories the division of labour between the available forms also differs, in that the learners display outspoken preferences for a restricted number of forms, whereas the native speakers draw on a wider pallet. The learners' fondness of particular general extenders appears to be motivated by a resemblance to a cognate form in their mother tongue (e.g. *or something, and so on*) and by the intensity of exposure to the form in the target language (e.g. *and stuff*). The

overrepresentation of disjunctive general extenders, and of certain functions of *or something (like that)* in particular, suggests that the learners use general extenders most prominently to deal with problems of language production sparking a greater need for markers of approximation. Although English has a wide range of pragmatic devices at the learners' disposal, they seem to fall back on a limited inventory of such markers.

Thus, the picture that emerges from our analysis confirms previous investigations insofar as, on the whole, the learners are not witnessed to elaborately deploy specific pragmalinguistic devices to foster interpersonal relations in conversation. Obviously the findings of this study cannot be generalised to all learners of English who are native speakers of Dutch, given the relatively small size of the corpus and the specific setting in which the conversations were conducted. It should be borne in mind, though, that the learners who participated in the study were overall advanced in English, were nearly in the final stage of the highest form of formal education in English available in their country, and are therefore expected to have reached near-native competence in English when they graduate from university. As much as the latter may be true for the lexical and grammatical levels, there is unmistakably still some way to go in the pragmatic domain. Further corpus investigations into various pragmatic elements (such as general extenders and pragmatic markers), in a variety of settings and for different types of language learners should inform both researchers and educationalists on the blind spots in language learners' pragmatic baggage, which in their turn could work some way towards raising pragmatic awareness among teachers and learners alike.

Appendix: Transcription Conventions

<Ir>	Interviewer turn starts
<Iee>	Interviewee turn starts
</Ir>	Interviewer turn ends
</Iee>	Interviewee turn ends
. . .	Empty pause: . (short), .. (medium), . . . (long)
[]	Phonetic annotations (e.g. the[i:], a[ei])
:	Vowel lengthening (e.g. *so:*)
=	Truncation
<X>	Unintelligible word
<overlap/>	Overlapping speech
<foreign>	Foreign words
<coughs>	Non-verbal sounds (e.g. coughing, laughing, sighing) and contextual comments (e.g. someone entering the room) are specified in angle brackets

References

Aijmer, K. (2002). *English discourse particles. Evidence from a corpus*. Amsterdam/Philadelphia: John Benjamins.

Aijmer, K. (2004). Pragmatic markers in spoken interlanguage. *Nordic Journal of English Studies, 3*, 173–190.

Aijmer, K. (2013). *Understanding pragmatic markers. A variational pragmatic approach*. Edinburgh: Edinburgh University Press.

Berkley, S. (2002). How to cure the verbal virus. A five-step treatment plan. *School for Champions*. http://www.school-for-champions.com/speaking/verbalvirus.htm. Accessed 23 July 2013.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Grammar of spoken and written English*. Harlow: Longman.

Brinton, L. J. (1996). *Pragmatic markers in English. Grammaticalization and discourse functions*. Berlin/New York: Mouton de Gruyter.

Buysse, L. (2011). The business of pragmatics. The case of discourse markers in the speech of students of Business English and English Linguistics. *ITL International Journal of Applied Linguistics, 161*, 10–30.

Buysse, L. (2012). *So* as a multifunctional discourse marker in native and learner speech. *Journal of Pragmatics, 44*, 1764–1782.

Buysse, L. (Forthcoming). 'Well it's economy and so on'. General extenders in learner and native English.

Callies, M. (2013). Advancing the research agenda of Interlanguage Pragmatics: The role of learner corpora. In J. Romero-Trillo (Ed.), *Yearbook of corpus linguistics and pragmatics 2013. New domains and methodologies* (pp. 9–36). Dordrecht: Springer.

Channell, J. (1994). *Vague language*. Oxford: Oxford University Press.

Cheshire, J. (2007). Discourse variation, grammaticalisation and stuff like that. *Journal of Sociolinguistics, 11*, 155–193.

De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *BELL. Belgian Journal of English Language and Literatures New Series, 2*, 225–246.

Dines, E. R. (1980). Variation in discourse – *and stuff like that*. *Language in Society, 9*, 13–31.

Dubois, S. (1992). Extension particles, etc. *Language Variation and Change, 4*, 179–203.

Evison, J., McCarthy, M., & O'Keeffe, A. (2007). 'Looking out for love and all the rest of it': Vague category markers as shared social space. In J. Cutting (Ed.), *Vague language explored* (pp. 138–157). Hampshire: Palgrave MacMillan.

Fernandez, J., & Yuldashev, A. (2011). Variation in the use of general extenders *and stuff* in instant messaging interaction. *Journal of Pragmatics, 43*, 2610–2626.

Gilquin, G. (2008). Hesitation markers among EFL learners: Pragmatic deficiency or difference? In J. Romero-Trillo (Ed.), *Pragmatics and corpus linguistics: A mutualistic entente* (pp. 119–149). Berlin: Walter de Gruyter.

Gilquin, G., De Cock, S., & Granger, S. (2010). *Louvain international database of spoken English interlanguage*. Louvain-la-Neuve: Presses Universitaires de Louvain.

Hasselgren, A. (2002). Learner corpora and language testing. Smallwords as markers of learner fluency. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 143–173). Amsterdam/Philadelphia: John Benjamins.

Hoffmann, S., Evert, S., Smith, N., Lee, D., & Prytz, Y. B. (2008). *Corpus linguistics with BNCweb: A practical guide*. Frankfurt am Main: Peter Lang.

Jucker, A. H. (1993). The discourse marker *well*: A relevance-theoretical account. *Journal of Pragmatics, 19*, 435–452.

Koester, A. (2007). 'About twelve thousand or so': Vagueness in North American and UK offices. In J. Cutting (Ed.), *Vague language explored* (pp. 40–61). Hampshire: Palgrave MacMillan.

Mukherjee, J. (2009). The grammar of conversation in advanced spoken learner English: Learner corpus data and language-pedagogical implications. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 203–230). Amsterdam/Philadelphia: John Benjamins.
Müller, S. (2005). *Discourse markers in native and non-native English discourse*. Amsterdam/ Philadelphia: John Benjamins.
Overstreet, M. (1999). *Whales, candlelight, and stuff like that: General extenders in English discourse*. Oxford: Oxford University Press.
Overstreet, M. (2005). And stuff *und so*: Investigating pragmatic expressions in English and German. *Journal of Pragmatics, 37*, 1845–1864.
Overstreet, M., & Yule, G. (1997). On being explicit and stuff in contemporary American English. *Journal of English Linguistics, 25*, 250–258.
Overstreet, M., & Yule, G. (2002). The metapragmatics of *and everything*. *Journal of Pragmatics, 34*, 785–794.
Palacios Martínez, I. M. (2011). "*I might, I might go I mean it depends on money things and stuff*". A preliminary analysis of general extenders in British teenagers' discourse. *Journal of Pragmatics, 43*, 2452–2470.
Parvaresh, V., Tavanger, M., Rasekh, A. E., & Izadi, D. (2012). About his friend, how good she is, *and this and that*: General extenders in native Persian and non-native English discourse. *Journal of Pragmatics, 44*, 261–279.
Romero-Trillo, J. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics, 34*, 769–784.
Scott, M. (2008). *WordSmith tools version 5*. Liverpool: Lexical Analysis Software.
Tagliamonte, S. A., & Denis, D. (2010). The stuff of change: General extenders in Toronto, Canada. *Journal of English Linguistics, 38*, 335–368.
Terraschke, A. (2007). Use of general extenders by German non-native speakers of English. *IRAL, 45*, 141–160.
Terraschke, A. (2010). *Or so, oder so, and stuff like that* – General extenders in New Zealand English, German and in learner language. *Intercultural Pragmatics, 7*, 449–469.
Terraschke, A., & Holmes, J. (2007). 'Und tralala': Vagueness and general extenders in German and New Zealand English. In J. Cutting (Ed.), *Vague language explored* (pp. 198–220). Hampshire: Palgrave MacMillan.
Ward, G., & Birner, B. (1993). The semantics and pragmatics of *and everything*. *Journal of Pragmatics, 19*, 205–214.

Oral Production of Discourse Markers by Intermediate Learners of Spanish: A Corpus Perspective

Leonardo Campillos Llanos and Paula González Gómez

Abstract This study describes the oral production of discourse markers by 40 (N = 40) learners of Spanish and compares it with usage by native speakers (N = 4). Our data belong to a learner corpus of oral interviews with university learners from over nine language backgrounds at intermediate level: A2 (N = 20) and B1 (N = 20) (*Common European Framework of Reference*). Semiautomatic part-of-speech (POS) tagging was used to count and retrieve the discourse markers produced by each group of learners and the group of native speakers. Results show a slight increase in the acquisition of these particles from A2 to B1, although the production is still lower when compared with the group of native speakers. Certain groups of students (especially Chinese learners) show a poorer usage of this category in our data, which could reveal a certain difficulty acquiring fluency at the discourse level. A breakdown of the most used discourse markers in our corpus (in native and non-native speakers, and at A2 and B1) is presented, as well as a distribution across interviews of the ten most frequent markers. Results are discussed comparing the usage data in our corpus with teaching guidelines for Spanish.

Keywords Spanish language • Second language acquisition • Learner corpus research • Discourse markers • Oral production

L. Campillos Llanos (✉)
Computational Linguistics Laboratory, Department of Linguistics and Modern Languages, Facultad de Filosofía y Letras, Universidad Autónoma de Madrid, C/Francisco Tomás y Valiente, 1 Cantoblanco Campus, 28049 Madrid, Spain
e-mail: leonardo.campillos@uam.es

P. González Gómez
Language Service, Universidad Autónoma de Madrid, C/Freud, 1, Cantoblanco Campus, 28049 Madrid, Spain

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2, DOI 10.1007/978-3-319-06007-1_11, © Springer International Publishing Switzerland 2014

1 Introduction

Interlanguage Pragmatics, as a subfield of cross-cultural pragmatics and second language (henceforth, L2) acquisition, is concerned with the study of L2 pragmatic use and L2 pragmatic learning/acquisition. Kasper (1992, 1996) points out that the study of use dominates Interlanguage Pragmatics and observes the need for more acquisitional studies to inform second language pedagogy.[1]

With this work we try to contribute to the research on Interlanguage Pragmatics by focusing on the acquisition of discourse markers (henceforth, DMs) in spoken L2 Spanish. Our study aims to analyse the non-native use of these devices by learners of Spanish at A2 and B1 levels (*Common European Framework of Reference*, henceforth *CEFR*; Council of Europe 2001) in order to address the following research questions:

- What markers are the most frequent in the oral discourse of learners with levels A2 and B1?
- What markers are the most frequent in the oral discourse of learners with different first language (henceforth, L1)?
- What differences exist in the use of DMs in native and non-native oral discourse?

The results of our data showed that there was a statistically significant difference between native and non-native speakers, who used fewer DMs at A2 and B1 levels. These outcomes support our claim that the appropriate use of DMs is crucial for the management of fluid and well-structured oral and written discourse. From a pedagogical point of view, these results lead us to suggest the explicit instruction of DMs to facilitate their acquisition, with the objective of helping students to develop their pragmatic-discourse competence.

This is a corpus-based study on foreign oral discourse and we are basically presenting a descriptive analysis, excluding variables such as the type of instruction received by learners. The reason is that, for some students, we do not know the nature of their instructional background, and to what degree it varies across learners. The work will start by explaining the theoretical framework and by introducing the research background on Interlanguage Pragmatics, especially on DMs. Then, we will focus on the existing literature regarding the acquisition of Spanish DMs, mainly corpus-based studies. After describing the methodology used,

[1] Research into acquisitional pragmatics should also be sociocognitive in orientation, just as we can put social and cognitive pragmatics together (Escandell 2004). A synthesis of the different aspects involved (learner and sociocultural context) configures the best framework for exploring how pragmatic competence is really acquired. Kasper and Rose (2002) offer a good review of the different theories of L2 pragmatic development (e.g. acculturation model, cognitive processing, and language socialization). Bardovi-Harlig (2013) also integrates social and cognitive factors in her consideration of the areas of research that are of interest for L2 pragmatics research (e.g. task design for the study of implicit and explicit knowledge, and the effect of environment on pragmatic development).

we will provide an analysis of the results. Finally, some pedagogical implications will be offered with the aim of improving the teaching of these particles in the classroom.

2 Theoretical Framework

The acquisition of pragmatic knowledge plays an essential role in the development of Communicative Competence (Hymes 1972; Canale and Swain 1980; Bachman 1990; Celce-Murcia et al. 1995). Pragmatic knowledge can be defined as the ability to use adequate language in different communicative situations. Consequences of pragmatic failure (Thomas 1983) can be worse at the social level than those produced by errors dealing with grammar or vocabulary. This mainly happens when the learner shows a high level of proficiency.

More research is still needed about pragmatic teachability. Nevertheless, in recent years it has been widely suggested that explicit pragmatic instruction improves the development of pragmatic competence (Kasper 1996; Rose and Kasper 2001; Jung 2002; Bardovi-Harlig and Mahan Taylor 2003). In order to guide instruction, we need to know how learners acquire pragmatic strategies and devices such as discourse markers. For that purpose, oral production data can provide very useful information about spoken interlanguage.

Some criticism has been made of the fact that studies in Interlanguage Pragmatics have paid more attention to the description of language use than to developmental aspects (Langer 2001), and also to the limited amount of research in Romance languages. Although it is true that research on the acquisition of the pragmatic level in Spanish is scarce, Galindo (2005) reported some references of studies regarding pragmatic transfer (none of them based on oral data). Other studies focused on speech acts and pragmatic errors in oral production (Blum-Kulka and House 1989; Koike 1989, 1996; Lorenzo-Dus 2001). These studies used listening comprehension tests or written questionnaires, which do not shed much light on the acquisition of oral skills. Nevertheless, by applying a Discourse approach, Lorenzo-Dus and Meara (2004) researched the assessment of spoken competence, and Sessarego (2009) examined the effects of instruction.

Non-native speakers have special difficulties mastering pragmatic devices that are more frequent or characteristic of speech, as is the case with discourse markers. These elements (also called *pragmatic markers* or *discourse particles*; see Blakemore 2004: 221; Aijmer and Simon-Vandenbergen 2011: 226) contribute significantly to the interpretation of the utterance due to its procedural meaning (Fraser 1999), as established in the Relevance theory (Blakemore 1987, 1992; Sperber and Wilson 1986).

A comprehensive and unified taxonomy of these particles is far from achieved. As Aijmer and Simon-Vandenbergen explain (2011: 228), Fraser (1996) distinguishes pragmatic markers, commentary markers, and parallel markers, but excluding words such as *well*. Further, some hesitators (e.g. *uhm, er*) are also sometimes

considered DMs (e.g. Gilquin 2008), but there is no general agreement on this criterion. However, some studies (e.g. Aijmer 2002) broadly identify a *textual function* and an *interpersonal function* of DMs.[2] Both functions are related to the notion of *indexicality* or *indexical function*, i.e. DMs denote the relation of an utterance to the immediate context, thereby creating cohesion (Fung and Carter 2007: 414; Aijmer and Simon-Vandenbergen 2011). While the *textual function* refers to the role these particles play in structuring the discourse, the *interpersonal function* concerns the expression of the speaker's attitudes or his/her role in the conversation. But regardless of the categorization of DMs, these devices should be present in the learners' interlanguage for them to successfully build their discourse, convey their attitude, or interact with other people in a natural way.

In the last three decades research on DMs has proliferated since the initial studies (e.g. Schiffrin 1987), and most research has been performed within a corpus-based approach, and in particular for English (e.g. Aijmer 2002; a full list of references is collected in O'Keeffe et al. 2007: 171–172). A recent methodology for the study of discourse markers has even proposed the use of translations to analyse the correspondences and differences between particles in the same semantic and pragmatic field (Aijmer and Simon-Vandenbergen 2003a, b, 2004; Simon-Vandenbergen and Willems 2011; see also the volume edited by Aijmer and Simon-Vandenbengen 2006).

Even though there has been less research devoted to the acquisition of these particles, the last decade has seen a growing interest in the application of corpus methods to address this question, especially in spoken interlanguage (e.g. Romero-Trillo 2002, 2004; Fuller 2003; Aijmer 2004, 2011; Müller 2005; Ramírez Verdugo and Romero-Trillo 2005; Llinares-García and Romero-Trillo 2006, 2008; Buysse 2007, 2011, 2012; Fung and Carter 2007; Gilquin 2008; Hellermann and Vergun 2007; Liao 2009; Mukherjee 2009; Polat 2011; Wei 2011; Zhao 2013; see the volumes edited by Romero-Trillo 2008, 2013).

3 Studies on Acquisition of Spanish Discourse Markers

Although there are already detailed studies on Spanish DMs (e.g. Briz 1993a, b; Casado Velarde 1998; Martín Zorraquino and Portolés 1999; Pons Bordería 2006), we are only aware of a few studies related to the acquisition of these particles. Fernández, in her error analysis of written texts (1990, 1997), included some DMs along with cohesive devices such as conjunctions, though other conversational markers were not covered (e.g. *bueno* or *hombre*). More recent research has gathered spoken data. For example, Díez Domínguez (2008) studied discourse

[2] Other studies such as Fung and Carter (2007) further distinguish a *structural function* and a *cognitive function* (e.g. to express hesitation or to indicate that the speaker is thinking).

markers by means of interviews with three learners who were studying in Spain and who had an advanced level (C1, *CEFR*).

Nogueira (2011) analysed the oral production of more than 20 Brazilian learners from upper-intermediate level (B2, *CEFR*), who performed a narrative task. The aim of Nogueira's study was to compare their use of DMs with those introduced in textbooks for learning Spanish. De la Fuente (2009) researched the influence of the type of instruction (explicit and implicit focus on form) in the acquisition of Spanish DMs by 24 English university learners. Hernández (2008, 2011) conducted similar research on the influence of explicit instruction and input flood on the acquisition of DMs by English learners of Spanish. Results from both studies suggested the positive influence of the explicit instruction on the acquisition of these markers. Finally, there is a study by Domenech (2008) on coordinate and subordinate connectors by Arabic learners (in oral and written tasks). Given that this study departed from a syntactic approach, we cannot strictly relate it to the pragmatic level.

A difficulty for research on discourse markers is the heterogeneity of methods used by each researcher in gathering the data: e.g. narrative tasks, interviews, consciousness raising tasks, reading aloud protocols, or questionnaires. This fact can make it awkward in comparing results among different studies.

Moreover, approaches for the analyses differ among investigators. While most studies are performed using a quantitative approach, other pieces of research are undertaken within a more descriptive or qualitative framework (Walsh 2013, for example, discusses this subject). It may be that the combination of both quantitative and qualitative approaches proves to be more beneficial than either one on its own.

Another hindrance for research in this area is the lack of a common taxonomy of DMs. Among the different classifications of Spanish DMs, we will mainly follow Martín Zorraquino and Portolés' classification (1999), a five-class typology that includes most of the DMs analysed in our study. In fact, it seems that discourse markers are sometimes considered a *wild card* category where different particles meet depending on the interests of each researcher (see, for example, the discussion by Fung and Carter 2007: 411; and by Aijmer and Simon-Vandenbergen 2011).

Thus, the research literature published on this topic should be read cautiously regarding the generalisation of results to different settings or type of learners.

4 Research Design

Our research is framed in the Learner Corpus Research (LCR) (Granger 2012). In particular, we followed a corpus-based approach (McEnery et al. 2006), since we used part-of-speech (henceforth, POS) annotation and frequency lists as research methods. As we have previously pointed out, corpus methodology is used more and more to tackle questions from the L2 acquisition research. Improvements in computing technology have made corpora progressively more available and easy to use. For English there already exist important learner corpus projects: e.g. the

International Corpus of Learner English (ICLE; Granger et al. 2009) and the Louvain International Database of Spoken English Interlanguage (LINDSEI; Gilquin et al. 2010). With regard to the Spanish language, there are at present resources such as the CEDEL corpus of written texts (Lozano and Mendikoetxea 2011). As for spoken skills, there is a longitudinal corpus of interviews (Díaz Rodríguez 2007), and the Spanish Learner Language Oral Corpus (SPLLOC; Mitchell et al. 2008).

While many research projects in the field of Spanish L2 have used both written and oral empirical data to test a hypothesis (Lafford and Salaberry 2003), corpus methods aim at developing a balanced and representative collection that could be re-employed for different topics of research. That is the reason why a key issue in the corpus paradigm is the design stage of the data collection (Granger 2008). Below, we will first describe the participants in our project; secondly, the design of the corpus; thirdly, the data elicitation methods; and finally, the procedure of the data analysis.

4.1 Participants and Corpus Design[3]

The participants were learners studying in Madrid thanks to international exchange programs (e.g. Erasmus), and almost all of them were between 19 and 26 years old, studying at an undergraduate or postgraduate level.[4] Our framework is limited to the use of Spanish in an academic context.

Unlike other research studies, our aim was to obtain spoken samples from several language backgrounds. A total of 40 (N = 40) subjects were interviewed, with four learners for every mother tongue. The participants were native speakers of nine different languages, which range from the Romance languages (Italian, French and Portuguese), the Germanic languages (English, German and Dutch) and the Slavic languages (Polish), to other Asian languages such as Chinese or Japanese. There was also another mixed group of four learners with other language backgrounds (one Finnish, one Korean, one Turkish and one Hungarian). Regarding the students' proficiency level, half of the learners (N = 20) were registered at the elementary level (A2), and the other half (N = 20) at the threshold level (B1) according to the *CEFR*.

Separately, a group of four native speakers (N = 4) with a similar educational level and age were interviewed with the same elicitation methods (control group).

[3] Further information about the corpus can be found at: http://cartago.lllf.uam.es/corele/index.html

[4] With respect to their learning background, some of the participants in our study had received formal instruction in Spanish in their countries, and others started in Madrid, combining formal instruction with natural-context acquisition. All of them had been in an immersion context for several months before being interviewed.

Table 1 Summary of data

Group	Level	N	DMs	% out of LUs	Lexical units	Mean	SD	Min	Max	Median	Range
Learners	A2	20	726	2.76	26,317	36.300	26.57	9	128	27	119
	B1	20	955	3.62	26,371	47.750	28.70	11	127	38	116
	Total:		**1,681**	**3.19**	52,688	**42.025**	27.91				
Native		4	**582**	**6.76**	8,610	**145.500****	69.73	82	241	129	159

N number of subjects, *DMs* frequency of discourse markers, *% out of LUs* ratio per lexical unit, *SD* standard deviation, *Min* minimum frequency, *Max* maximum frequency
**Results are significant at $p = 0.0031$

These interviews were used as a benchmark to compare the use of DMs between native and non-native speakers and to unveil certain phenomena that were more frequent in the spoken discourse in both groups.

Each interview lasted about 15–20 min, so approximately more than 1 h was recorded for every group of participants. The recordings were made at the Universidad Complutense and the Universidad Autónoma de Madrid during the academic courses 2007–2008, 2008–2009, and 2009–2010. On the whole, more than 13 h were collected for the learner corpus, which comprised a total of over 55,000 tokens, whereas the control group exceeded 9,000 tokens (in this count, a token was considered as the word between two white spaces, so the discourse marker *es decir,* 'that is to say', counted as two tokens). With a view to normalising the data and achieving more precision when analysing the lexis, we performed another count by lexical unit (with multiwords such as *por otra parte,* 'on the other hand', or *o sea,* 'that is', counting as one unit). Following these criteria, the learner corpus amounted to 52,688 lexical units, and the native speakers' group, to 8,610 (Table 1).

4.2 Elicitation Methods

The interview, carried out by the researcher for the learners and the native speakers, was voluntary and, learners had the opportunity to receive an explanation of their errors after its completion. The interview began with a brief introduction (e.g. providing information such as their studies, years studying Spanish, time in a Spanish-speaking country). Subsequently, in order to obtain comparable data, they all accomplished the same tasks, which were similar to those in foreign language examinations:

- Providing a description of two photographs (in this project, related to food).
- Retelling a story from pictures, to make learners use the past tense and to ask them a question involving two speech acts (a suggestion and a request).

The last part of the interview was more spontaneous, and the researcher asked the learners to give opinions about non-personal topics (in our interview, related to

food; e.g. changes in food habits in present-day society, or differences between food in Spain and that from the learner's country).

4.3 Procedure and Data Analysis

After the recording, interviews were manually transcribed, and the transcriptions were subsequently POS-tagged with a view to studying and describing the learners' spoken production of the different morphological categories. POS tagging was partially automated with GRAMPAL, a morphological processor of Spanish that was adapted for spoken data (Moreno and Guirao 2006). GRAMPAL handles multiword units (such as *gracias a,* 'thanks to'), and it is also able to tag discourse markers, which are often multiword units: e.g. *por otro lado* ('on the other hand'), *quiero decir* ('I mean'), or *por eso* ('because of that').

A manual revision of the tagging was performed so as to correct ambiguities concerning the automatic assignment of categories. Some ambiguities were due to homonymy: e.g. *vale,* which can be a noun ('voucher') or a discourse marker ('OK'). Other ambiguities were due to incorrect categorization: e.g. *bueno*, which can be an adjective in *hombre bueno* ('good man') or a discourse marker in *¡Hombre! Bueno*... ('Hey! Well...'). Finally, other ambiguities arose in phrase chunking: e.g. *es decir* ('that is'), which is not a discourse marker in *Lo que hace es decir tonterías* ('What he is doing is saying silly things').

After the POS-tagging process, a list of DMs was obtained for each group of learners, as well as for the group of native speakers. In this count, we did not take into consideration markers from other languages that were found in learner's discourse as a result of interference phenomena.[5] For example, in the following utterance, the conversational marker *bem* (from Portuguese) was used instead of *bueno,* 'well' (we show a simplified version of the transcriptions):

JUS: ***bom*** *yo creo que esta é una paella* (PORWA2_2)
'well I think that this is a paella'

Among the learners from the A2 level, nine DMs interfered while they were speaking in Spanish: *bom* (twice), *like* ('como', twice), *por exemplo* ('for example', twice), *ben* ('right', instead of Spanish *bien*, once) and *é que* ('it is just that', once). Among the learners at the B1 level, this interference decreased slightly, since only four non-Spanish DMs were registered: *like* (two times), *I mean* ('quiero decir'), and *bon* ('right').

[5] We decided to include *OK* in our study because it is used more frequently with the same meaning as 'de acuerdo' or 'vale' by native speakers, due to the influence of English. Notwithstanding that fact, *OK* is not found among the twenty-five most frequent discourse markers used by native speakers (Table 3).

Similarly, we did not include in this count misformations due to interference or false hypotheses about the right form. For example, in the following utterance the right connector should have been *por otro lado*:

EVE: creo que ***en el otro lado*** *es un costumbre bueno* (DUTWB1)
‘on the other hand, I think that it is a good habit’

Only six non-existent DMs were documented in our data (three at the A2 level and three at the B1 level).

5 Results

Considering the above-mentioned criteria, a total of 1,681 DMs were registered in the learner group. The learners at the A2 level produced 726 DMs, with a mean of 36.30 markers and about 2.76 markers for every one hundred lexical units. At the B1 level this amount slightly increased to 955 in our data, with a mean of 47.75 markers per interview and approximately 3.62 every one hundred lexical units (Fig. 1 and Table 1). In turn, the native speakers produced 582 DMs, with a higher mean per interview (145.50) when compared to that of the learners’, and a higher ratio of DMs per lexical unit (6.76 %).

The native speakers’ production of DMs varied among the four participants interviewed, as the standard deviation measure and the range was greater than that of the groups of learners. This would imply that the mean value of DMs produced by the native speakers was not as reliable as the mean of the learners. In order to gain further insight into the difference in the use of discourse markers, we performed a statistical test comparing their frequency among the three groups. As the distribution of the data did not pass any test of normality (Kolmogorov-

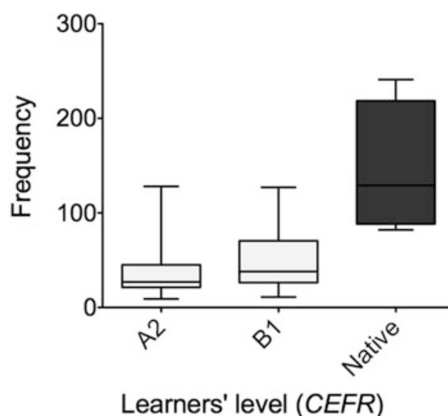


Fig. 1 Distribution of markers across levels and comparison with native speakers. The *outer lines* of the plot represent the maximum and minimum values in each group

Table 2 Summary of data

Group	L1	N	Disc. mark.	% out of LUs	Lexical units	Mean	SD	Min	Max	Range
Learners	Portuguese	4	194	2.64	7,329	48.500	22.43	25	71	46
	Italian	4	195	3.18	6,133	48.750	31.56	21	79	58
	French	4	349	5.44	6,413	87.250	46.54	44	128	84
	English	4	186	3.93	4,731	46.500	27.29	11	77	66
	German	4	133	2.91	4,573	33.250	15.96	17	55	38
	Dutch	4	153	2.95	5,186	38.250	12.12	26	51	25
	Polish	4	109	2.06	5,300	27.250	8.61	22	40	18
	Chinese	4	72	2.00	3,592	18.000	5.47	14	26	12
	Japanese	4	163	3.49	4,664	40.750	27.74	11	78	67
	Other	4	127	2.66	4,767	31.750	15.69	9	45	36
	Total	40	**1,681**	**3.19**	52,688	**42.025**	27.89	9	128	118
Native		4	**582**	**6.76**	8,610	**145.500**	69.92	82	241	159

L1 learners' mother tongue, *N* number of subjects, *Disc. mark.*, frequency of discourse markers, *% out of LUs* ratio of discourse marker per lexical unit (in each group), *SD* standard deviation, *Min* minimum frequency, *Max* maximum frequency

Smirnoff and D'Agostino), a non-parametric test was chosen. Notwithstanding the aforementioned variance, significant differences among the A2, the B1 and the group of native speakers were revealed (Kruskal-Wallis: $H = 11.57$, 2df, $p = 0.0031$). In contrast, we did not find significant differences between the A2 and the B1 group (Mann-Whitney, one-tailed: $p = 0.0648$).

With regard to the learner's L1, we observed that Chinese learners in our data produced the lowest number of discourse markers (with a mean of 18 per interview, and a ratio of two for every one hundred lexical units; Table 2). On the contrary, the French learners were the group that used more DMs (a mean of 87.25, and a ratio of 5.44 %). Nevertheless, the variance within the French group was also quite high (SD = 46.54, with a minimum of 44 DMs in one interview but a maximum of 128 in another). Due to the scarce number of participants in each L1 group, we did not perform further statistical analyses.

The difference between the Chinese group and the learners whose L1 is closer to Spanish (such as French, Portuguese, and Italian) could reveal certain difficulties in discourse and pragmatic skills in relation to the Spanish language. However, the typological explanation is not coherent with the high frequency of discourse markers used by other learners whose L1 is not Indo-European. An example in our data is Japanese learners, who produced a mean of 40.75 per interview (with a ratio of 3.49 %), even higher than that of the native speakers of Polish (an Indo-European language), who showed a mean of 8.61 markers (a ratio of 2.06 %).

The differences between the equivalence of discourse markers in the L1 and the L2 could better account for some of these results. For example, some researchers (Liao 2009: 1320) have confirmed the underuse of certain English discourse markers by Chinese learners due to the lack of a similar counterpart in Mandarin

(e.g. *well, I mean*).[6] This aspect deserves to be studied further by analysing the pragmatic and semantic correspondence of the markers in each language through Aijmer and Simon-Vandenbergen's model (2004).

In addition, extralinguistic factors could also explain these results. Apart from the fact that Chinese participants in our study showed a low fluency during the interview, teaching-related factors can possibly justify the underuse of these particles. Teaching methodologies of English or French language have been including the explicit instruction of discourse markers for the last decade. Interestingly, learners whose L1 is French or English showed a high rate of DMs. We are inclined to think that these learners acquired a good command of these devices when they learnt a language such as English or French, and they benefited from this experience when learning an L3 (e.g. Spanish), unlike Chinese students. Other variables such as the learning background of Spanish (e.g. in an academic setting or in a study abroad program), the time of exposure to the L2, or the discourse abilities of each learner (his/her idiolect) should also be taken into account in future studies. Given the scarce number of participants in our study, conclusions cannot be generalised.

Regarding the frequency of use of each discourse marker, Table 3 shows a breakdown of the 25 most frequent particles in our study (grouped by learners and native speakers). Among the 25 most frequent DMs, 16 were produced by both native and non-native participants. Most of these markers belong to the following types according to Martín Zorraquino and Portoles' typology (1999): connectors (*pues, entonces*), reformulators (*o sea*), argumentative operators (*por ejemplo*), and conversational markers (*bueno, vale, vamos, claro, oye, por favor, mira, oye, ¿no?*). The other markers produced by both learners and native speakers (*como, es que, por eso, a ver*) are not included in that typology. It is interesting that there are no information structurers among them (only the learners used *primero*, 'first', which introduces a set of points or ideas).

Nevertheless, the learners did not produce some conversational markers (*nada, hombre, vamos, efectivamente*), certain contra-argumentative connectors (*sin embargo*), specific additive connectors (*y demás, y tal, etcétera*) or reformulators (*o sea que,* though they used *o sea*). These results could be explained by the elicitation tasks employed in the study. These tasks were mainly of descriptive and narrative nature and did not favour the different functions associated with the mentioned markers. Some particles (e.g. *hombre* or *nada*[7]) are frequently used by

[6] Yao (2012) compared Spanish and Chinese discourse markers and explained some differences among them. In Chinese, modal particles do not convey a conceptual meaning, and, unlike Spanish, these elements can be used as discourse markers. In addition, most discourse markers in Chinese appear at the beginning of the utterance. In Spanish, discourse markers can occur either at the beginning, the middle or the end of the utterance. Further information on Chinese DMs can be found in Liao (1986), Miracle (1991), and Liu (2009).

[7] These conversational markers have received little attention in studies regarding their description or teaching (Gozalo 2008; Gaviño 2011; Schmer 2012).

Table 3 The twenty-five most frequent markers (Learners and native speakers)

Learners (N = 40)			Native speakers (N = 4)		
Discourse marker	Freq.	Rel. freq. (%)	Discourse marker	Freq.	Rel. freq. (%)
Vale	283	16.84	*Pues*	193	33.16
Como	192	11.42	*Bueno*	104	17.87
Pues	141	8.39	*Es que*	33	5.67
Es que	135	8.03	*O sea*	28	4.81
Bueno	133	7.91	*¿No?*	27	4.64
Entonces	117	6.96	*Claro*	26	4.47
Por eso	117	6.96	*Vale*	25	4.30
Ok	95	5.65	*Entonces*	20	3.44
Claro	94	5.59	*Pero bueno*	15	2.58
¿No?	72	4.28	*Por ejemplo*	13	2.23
Por ejemplo	60	3.57	*Como*	11	1.89
Por favor	51	3.03	*A ver*	9	1.55
¿Sabes?	37	2.20	*Hombre*	6	1.03
O sea	27	1.61	*Y tal*	6	1.03
Es como	22	1.31	*Efectivamente*	4	0.69
Bien	21	1.25	*Etcétera*	4	0.69
A ver	10	0.59	*Mira*	4	0.69
Por supuesto	10	0.59	*Por favor*	4	0.69
Primero	9	0.54	*Vamos*	4	0.69
En fin	5	0.30	*Y demás*	4	0.69
Mira	5	0.30	*Nada*	3	0.52
Luego	4	0.24	*O sea que*	3	0.52
Oye	3	0.18	*Oye*	3	0.52
Pero bueno	3	0.18	*Por eso*	3	0.52
Quiero decir	3	0.18	*Sin embargo*	3	0.52

native speakers and therefore present in the input received by learners. However, these markers are not usually found in pedagogic materials or considered in formal instruction.

If we consider the 20 most frequent markers used by the learners (see Table 4), production was similar at both the A2 and B1 levels. Additionally, comparing Tables 3 and 4, we observed that the B1 learners produced seven of the ten most frequent markers in the native group (*pues, bueno, es que, ¿no?, claro, vale, entonces*), while the A2 learners just five (*pues, bueno, es que, claro, entonces*).

Table 5 breaks down the distribution across interviews of the ten most frequent discourse markers. According to our data (Fig. 2), when compared to the A2 group, we identified a spread of use of almost every marker in the interviews of the B1 group. The only exceptions were the following: *OK* and *por eso* (with a lower distribution), and *entonces* and *pues* (with the same distribution). These results suggest a positive progress in the acquisition of DMs.

Table 4 The twenty most frequent markers used by learners (A2 and B1 level)

A2 (N = 20)			B1 (N = 20)		
Discourse marker	Freq.	Rel. freq. (%)	Discourse marker	Freq.	Rel. freq. (%)
Vale	130	17.91	*Vale*	153	16.02
Pues	100	13.77	*Como*	115	12.04
Como	77	10.61	*Bueno*	111	11.62
Entonces	62	8.54	*Es que*	74	7.75
Es que	61	8.40	*Entonces*	55	5.76
Por eso	53	7.30	*¿No?*	52	5.45
Ok	50	6.89	*Claro*	50	5.24
Claro	44	6.06	*Ok*	45	4.71
Por ejemplo	25	3.44	*Pues*	41	4.29
Por favor	24	3.31	*Por eso*	64	6.70
Bueno	22	3.03	*Por ejemplo*	35	3.66
¿No?	20	2.75	*O sea*	25	2.62
Es como	17	2.34	*Por favor*	27	2.83
¿Sabes?	16	2.20	*¿Sabes?*	21	2.20
Por supuesto	6	0.83	*Bien*	18	1.88
Bien	3	0.41	*A ver*	10	1.05
O sea	2	0.28	*Primero*	9	0.94
Pero bueno	2	0.28	*Es como*	5	0.52
Por otra parte	2	0.28	*En fin*	4	0.42
En fin	1	0.14	*Luego*	4	0.42

Table 5 Distribution across interviews from each level (*CFER*) of the ten most frequent discourse markers produced by the learners in our study

Discourse	A2 (N = 20)		B1 (N = 20)	
markers	Distribution	Frequency (% out of 20)	Distribution	Frequency (% out of 20)
Vale	17	85	19	95
Como	12	60	15	75
Entonces	12	60	12	60
Claro	11	55	12	60
¿No?	10	50	11	55
Por eso	10	50	9	45
Es que	9	45	11	55
Pues	9	45	9	45
OK	8	40	4	20
Bueno	3	15	6	30

6 Discussion and Pedagogical Implications

The types of markers produced by the learners and native speakers were very similar. The number of information structurers or contra-argumentative connectors was scarce, and the most frequent were conversational markers (*vale, bueno, ¿no?, claro*) and consecutive connectors (*por eso, pues, entonces*). This could be

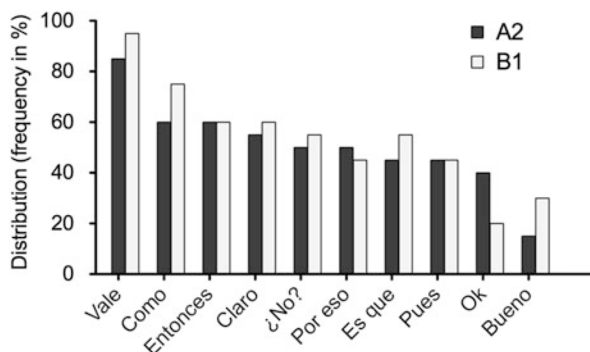


Fig. 2 Distribution across interviews (expressed as percentage) of the ten most frequent discourse markers used by learners at A2 and B1

explained by the elicitation method and also by the natural context of L2 acquisition: markers that are frequently used by native speakers configure the input received by learners.

In the distribution across interviews we observed that, among the ten most frequent markers, there were mainly connectors (*entonces, por eso, pues*) and conversational markers (*vale, claro, no, bueno, OK*). As before, we also found in these data two markers that have received less attention in theoretical studies. An example is *como,* which has traditionally been considered a hedge or an expletive. Another example is *es que,* which some scholars classify as an argumentative connector that introduces an explanation or a justification (Fernández Leborans 1992; Pons Bordería 1998; Porroche Ballesteros 1998; Santos Río 2003). These markers are not included in the classification offered in Martín Zorraquino and Portolés (1999).

As suggested above, some interesting trends in the use of discourse markers were observed when comparing both proficiency levels. For example, the conversational marker *bueno* was not one of the ten most frequent markers at level A2, but it was the third most frequent marker at level B1 (Table 4). Its use by the learners seemed to be very close to the native production in which it was the second most frequent marker (Table 3). If we consider its distribution (Table 5), *bueno* was used in six interviews at level B1, twice the rate used at level A2. Curiously, the occurrence of *bueno* was markedly reduced in the learners' production when it appeared combined with the conjunction *pero*. According to Table 3, *pero bueno* was registered 15 times among the native speakers, in contrast to only three instances among the learners.

The case of the marker *OK* is also interesting. Its occurrence in the groups of learners was greater at the A2 level and by speakers whose L1 was English. The interlinguistic interference could explain these data. At the B1 level, its use was reduced in favor of the Spanish equivalent *vale*.

The distribution of markers such as *pues, entonces,* and *por eso* across participants did not only increase (in the case of the first two particles), but was reduced at B1 (Table 5). This fact could be attributed to the learner's idiolect (especially in the case of *por eso*). Some students showed a tendency to employ some markers with an exclusively expletive function.

Table 6 Discourse markers proposed in the *PCIC* registered among the most frequent in our data

Level	Discourse markers (*PCIC*)
A1	*Por ejemplo, ¿no?*
A2	*Por eso, entonces, primero, mira*
B1	*Como, es que, pues, o sea, claro, ¿sabes?*
B2	*Por supuesto, en fin, bueno*

However, although the statistical test did not yield significant results when comparing the two levels (Table 1), frequency and distribution analyses indicated a slightly progressive increase in the production of markers from the A2 level to the B1 level.

Following this, we can compare our data with the established guidelines for the teaching of Spanish as a second language (*Plan Curricular del Instituto Cervantes,* from here on, *PCIC*; Instituto Cervantes 2006) (Table 6). According to this reference document, the marker *¿no?* should be introduced at A1, *oye*/*oiga* and *mira*/*mire* at A2, and *¿sabes?* at B1. Indeed, learners in our data did use them at these levels (excepting *oye*/*oiga*). Further, they did use most of the markers for the A1-B1 levels, despite not having produced instances of some particles established for B1. The non-native speakers even used some particles corresponding to B2.

To sum up, in general, learners in our study produced a wide range of markers established in the guidelines for Spanish language teaching. Nevertheless certain DMs, which were found in the native oral discourse, were not produced by the learners. Therefore we consider that more effort should be made in the teaching of those markers. Pragmatics is teachable and a majority of studies favor explicit approaches to the teaching of L2 pragmatics (Koike 1989; Bouton 1999; House 1996; Kasper 1997; Kasper and Rose 2002), and particularly of DMs (Yoshimi 2001; Hernández 2008, 2011; Hernández and Rodríguez-González 2012). Explicit teaching is especially important at low proficiency levels for which pedagogic materials offer exclusively implicit teaching, and in foreign language environment, because there are fewer learning opportunities outside the classroom.

The instruction can be carried out from the lowest levels of learning (Kasper 1997) and we suggest following the three I's methodology: Illustration, Interaction and Induction (Carter and McCarthy 1995). This methodology involves basic principles such as the exposition to authentic input, the activation of acquired knowledge, and consciousness-raising procedures. The three stages should be complemented with communicative practice and corrective feedback in order to improve the effectiveness of the teaching of DMs.

7 Concluding Remarks

In this work we have presented a corpus-based description of the use of discourse markers in native and non-native oral production at intermediate levels. Results of the analysis can be summarised as follows. With regard to our first research question—i.e. what markers are the most frequent in the oral discourse of learners at A2 and B1

levels?—we found a wide range of Spanish DMs in the oral production of learners at these levels of proficiency. However, the statistical tests did not reveal significant results when comparing the A2 and B1 groups. Even so, a progress in the acquisition was observed. Not only the number of markers increased from A2 to B1, but also the distribution of markers across participants spread as well.

Regarding the second research question—i.e. what markers are the most frequent in the oral discourse of learners with different L1?—we observed dissimilarities in the production of these particles, both in number and type of marker. The explanation of these results may be to a certain extent interlinguistic. For example, the equivalence of markers in learners' first language can influence the usage of these particles, as we found among our Chinese learners. Additionally, other factors such as the type of instruction received by each learner or his/her learning context (e.g. in an academic setting or in a study abroad program) may also be involved in the learners' use and acquisition of DMs.

As for the third research question—i.e. what differences exist in the use of discourse markers in native and non-native discourse?—the statistical tests yielded significant results when comparing the native and non-native speakers groups (the frequency of markers used by the learners is significantly lower). The types of markers used by the learners were also similar to those used by the native speakers, probably due to the elicitation task. In addition, the learners produced most of the markers proposed by the teaching guidelines for Spanish language established in the *PCIC* for the A2 and B1 levels.

Further analyses with data from more participants could confirm these results. As well, further research should be carried out regarding qualitative aspects of production. An analysis of the usage context of each marker is necessary to check if learners use these particles accurately, and what functions pose a problem in the acquisition of polyfunctional markers (e.g. *bueno* or *pues*). Likewise, the semantic and pragmatic differences between equivalent discourse markers in each language deserve to be closely studied following the model proposed by Aijmer and Simon-Vandenberger (2004). These data, along with those from further studies on the degree of influence of the explicit teaching on the acquisition of discourse markers, will certainly help guide the explicit teaching of these devices.

Acknowledgments This work was supported by the Regional Government of Madrid and the European Social Fund under a PhD grant; and the Comisión Interministerial de Ciencia y Tecnología (CICYT) [project TIN2010-20644-C03-03]. The authors gratefully acknowledge the anonymous reviewers for their valuable comments, and would like to thank T. Shipley and C. Herrero for their assistance in the linguistic revision of the manuscript.

References

Aijmer, K. (2002). *English discourse particles: Evidence from a corpus*. Amsterdam/Philadelphia: John Benjamins.

Aijmer, K. (2004). Pragmatic markers in spoken interlanguage. *Nordic Journal of English Studies, 3*(1), 173–190.

Aijmer, K. (2011). *Well I'm not sure I think*... The use of *well* by non-native speakers. *International Journal of Corpus Linguistics, 16*(2), 231–254.

Aijmer, K., & Simon-Vandenbergen, A.-M. (2003a). The discourse particle *well* and its equivalents in Swedish and Dutch. *Linguistics, 41*(6), 1123–1161.

Aijmer, K., & Simon-Vandenbergen, A.-M. (2003b). The expectation marker *of course* in a cross-linguistic perspective. *Languages in Contrast, 31*, 13–43.

Aijmer, K., & Simon-Vandenbergen, A.-M. (2004). A model and a methodology for the study of pragmatic markers: The semantic field of expectation. *Journal of Pragmatics, 36*, 1781–1805.

Aijmer, K., & Simon-Vandenbengen, A.-M. (Eds.). (2006). *Pragmatic markers in contrast*. Oxford: Elsevier.

Aijmer, K., & Simon-Vandenbergen, A.-M. (2011). Pragmatic markers. In J. Zienkowski, J.-O. Östman, & J. Verschueren (Eds.), *Discursive pragmatics* (Handbook of Pragmatic Highlights, Vol. 8, pp. 223–247). Amsterdam/Philadelphia: John Benjamins.

Bachman, L. (1990). Habilidad lingüística comunicativa. In M. Llobera (Ed.), *Competencia comunicativa* (pp. 105–129, 1995). Madrid: Edelsa.

Bardovi-Harlig, K. (2013). Developing L2 pragmatics. *Language Learning, 63*(1), 68–86.

Bardovi-Harlig, K., & Mahan Taylor, R. (Eds.). (2003). *Teaching pragmatics*. Washington, DC: United States Department of State.

Blakemore, D. (1987). *Semantic constraints on relevance*. Oxford: Blackwell.

Blakemore, D. (1992). *Understanding utterances*. Oxford: Blackwell.

Blakemore, D. (2004). Discourse markers. In L. R. Horn & G. Ward (Eds.), *Handbook of pragmatics* (Blackwell handbooks in linguistics, pp. 221–240). London: Blackwell.

Blum-Kulka, S., & House, J. (1989). Cross cultural and situational variation in requesting behavior. In S. Blum-Kulka, J. House, & G. Kasper (Eds.), *Cross-cultural pragmatics: Requests and apologies* (pp. 123–154). Norwood: Ablex.

Bouton, L. (1999). Developing non-native speaker skills in interpreting conversational implicatures in English: Explicit teaching can ease the process. In E. Hinkel (Ed.), *Culture in second language teaching and learning* (pp. 47–70). Cambridge: Cambridge University Press.

Briz, A. (1993a). Los conectores pragmáticos en español coloquial. I: Su papel argumentativo. *Contextos, 11*, 145–188.

Briz, A. (1993b). Los conectores pragmáticos en la conversación coloquial. II. Su papel metadiscursivo. *Español actual, 59*, 39–56.

Buysse, L. (2007). Discourse marker *so* in the English of Flemish university students. *Belgian Journal of English Language and Literatures,* New Series *5*: 79–95.

Buysse, L. (2011). The business of pragmatics: The case of discourse markers in the speech of students of business English and English linguistics. *ITL: International Journal of Applied Linguistics, 161*, 10–30.

Buysse, L. (2012). *So* as a multifunctional discourse marker in native and learner speech. *Journal of Pragmatics, 44*, 1764–1782.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1–47.

Carter, R., & McCarthy, M. (1995). Grammar and the spoken language. *Applied Linguistics, 16*(2), 141–158.

Casado Velarde, M. (1998). Lingüística del texto y marcadores del discurso. In M. A. Martín Zorraquino & E. Montolío Durán (Eds.), *Marcadores del discurso. Teoría y análisis* (pp. 55–70). Madrid: Arco/Libros, S.L.

Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). A pedagogical framework for communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics, 6*(2), 5–35.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

De la Fuente, M. J. (2009). The role of pedagogical tasks and focus on form on acquisition of discourse markers by advanced language learners. In R. P. Leow, H. Campos, & D. Lardiere

(Eds.), *Little words: Their history, phonology, syntax, semantics, pragmatics and acquisition* (pp. 211–221). Washington, DC: Georgetown University Press.
Díaz Rodríguez, L. (2007). *Interlengua española: estudio de casos*. Barcelona: Printulibro Intergrup.
Díez Domínguez, P. (2008). *Análisis de los marcadores discursivos más usados en el habla de tres estudiantes extranjeros de E/LE nivel C1 tras una estancia de un mes en España*. Master's thesis, University of Barcelona, Barcelona.
Domenech Val, A. (2008). *La conectividad en español no nativo: un estudio sobre hablantes arabófonos*. Master's thesis, University of Barcelona, Barcelona.
Escandell, V. (2004). Norms and principles. Putting social and cognitive pragmatics together. In R. Márquez-Reiter & M. E. Placencia (Eds.), *Current trends in the pragmatics of Spanish* (pp. 347–371). Amsterdam/Philadelphia: John Benjamins.
Fernández Leborans, M. J. (1992). La oración del tipo 'es que...'. *Verba, 19*, 223–239.
Fernández López, S. (1990). Problemas discursivos en la interlengua de aprendices de español como lengua extranjera. In S. Montesa Peydró, & A. M. Garrido (Eds.), *Proceedings of the second national ASELE conference* (pp. 267–278, 1994). Málaga: ASELE.
Fernández López, S. (1997). *Interlengua y análisis de errores en el aprendizaje del español*. Madrid: Edelsa.
Fraser, B. (1996). Pragmatic markers. *Pragmatics, 6*(2), 167–190.
Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics, 31*, 931–952.
Fuller, J. M. (2003). Discourse marker use across speech contexts: A comparison of native and non-native speaker performance. *Multilingua, 22*, 185–208.
Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics, 28*(3), 410–439.
Galindo Merino, M. (2005). La transferencia pragmática en el aprendizaje de ELE. In A. Álvarez et al. (Eds.), *Proceedings of the XVI international ASELE conference* (pp. 289–297). Oviedo: Ediciones de la Universidad de Oviedo
Gaviño, V. (2011). Operaciones metalingüísticas del marcador discursive *hombre*. *Marcoele, 12*. http://marcoele.com/descargas/12/gavino-marcador_hombre.pdf. Accessed 9 Aug 2013.
Gilquin, G. (2008). Hesitation markers among EFL learners: Pragmatic deficiency or difference? In J. Romero-Trillo (Ed.), *Pragmatics and corpus linguistics. A mutualistic entente* (Mouton Series in Pragmatics, Vol. 2, pp. 119–149). Berlin: Mouton de Gruyter.
Gilquin, G., De Cock, S., & Granger, S. (2010). *The Louvain international database of spoken English interlanguage*. Louvain-la-Neuve: Presses Universitaires de Louvain.
Gozalo, P. (2008). El marcador discursivo hombre y su tratamiento en el aula de E/LE. *REDELE, 14*. www.doredin.mec.es/documentos/00820083000393.pdf. Accessed 3 Sept 2013.
Granger, S. (2008). Learner corpora. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (Vol. 1, pp. 259–275). Berlin: Walter de Gruyter.
Granger, S. (2012). How to use foreign and second language learner corpora. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 7–29). London: Blackwell.
Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). *International corpus of learner English V2*. Louvain-la-Neuve: Presses universitaires de Louvain.
Hellermann, J., & Vergun, A. (2007). Language which is not taught: The discourse marker use of beginning adult learners of English. *Journal of Pragmatics, 39*(1), 157–179.
Hernández, T. A. (2008). The effect of explicit instruction and input flood on students' use of Spanish discourse markers on a simulated oral proficiency interview. *Hispania, 91*, 665–675.
Hernández, T. A. (2011). Re-examining the role of explicit instruction and input flood on the acquisition of Spanish discourse markers. *Language Teaching Research, 15*(2), 159–182.
Hernández, T. A., & Rodríguez-González, E. (2012). Impact of instruction on the use of L2 discourse markers. *Journal of Second Language Teaching and Research, 2*(1), 3–31.
House, J. (1996). Developing pragmatic fluency in English as a foreign language: Routines and metapragmatic awareness. *Studies in Second Language Acquisition, 18*(2), 225–252.
Hymes, D. (1972). Acerca de la competencia comunicativa. In M. Llobera (Ed.), *Competencia comunicativa* (pp. 27–47, 1995). Madrid: Edelsa.

Instituto Cervantes. (2006). *Plan Curricular del Instituto Cervantes*. http://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/. Accessed 3 Sept 2013.

Jung, J. Y. (2002). Issues in acquisitional pragmatics. *Working Papers in TESOL & Applied Linguistics, 2*(3), 1–13.

Kasper, G. (1992). Pragmatic transfer. *Second Language Research, 8*(3), 203–231.

Kasper, G. (1996). Interlanguage pragmatics in SLA (Introduction to thematic issue). *Studies in Second Language Acquisition, 18*(2), 145–148.

Kasper, G. (1997). *Can pragmatic competence be taught?* Honolulu: University of Hawai'i, Second Language Teaching & Curriculum Center. www.nflrc.hawaii.edu/NetWorks/NW6/. Accessed 3 Sept 2013.

Kasper, G., & Rose, K. (2002). *Pragmatic development in a second language*. Malden: Blackwell.

Koike, D. A. (1989). Pragmatic competence and adult L2 acquisition: Speech acts in interlanguage. *The Modern Language Journal, 73*, 279–289.

Koike, D. A. (1996). Transfer of pragmatic competence and suggestions in Spanish foreign language learning. In S. M. Gass & J. Neu (Eds.), *Speech acts across cultures: Challenges to communication in a second language* (pp. 257–281). Berlin: Mouton de Gruyter.

Lafford, B. A., & Salaberry, M. R. (2003). *Spanish second language acquisition: State of the science*. Washington, DC: Georgetown University Press.

Langer, B. (2001). Teaching pragmatics forms in Spanish. *Segundas lenguas e inmigración en red, 5*, 5–34.

Liao, Q. (1986). Discourse connectives in modern Mandarin Chinese. *Zhongguo Yuwen, 6*, 413–427.

Liao, S. (2009). Variation in the use of discourse markers by Chinese teaching assistants in the US. *Journal of Pragmatics, 41*, 1313–1328.

Liu, B. (2009). Chinese discourse markers in oral speech of mainland Mandarin speakers. In Y. Xiao (Ed.), *Proceedings of the 21st North American Conference on Chinese Linguistics (NACCL-21)* (Vol. 2, pp. 358–374). Smithfield: Bryant University. Available at: https://naccl.osu.edu/sites/naccl.osu.edu/files/NACCL-21_Vol._2–Binmei_Liu–pp._358-374.pdf. Accessed 27 Aug 2013.

Llinares-García, A., & Romero-Trillo, J. (2006). Discourse markers in the EFL classroom. In M. Carretero, L. Hidalgo Downing, J. Lavid, E. Martínez-Caro, J. Neff, S. Pérez de Ayala, & E. Sanchez-Pardo (Eds.), *A pleasure of life in words: A festschrift for Angela Downing* (Vol. 2, pp. 163–176). Madrid: Complutense University.

Llinares-García, A., & Romero-Trillo, J. (2008). The pragmatic role of discourse markers of native and non-native teachers in CLIL contexts. In J. Romero-Trillo (Ed.), *Pragmatics and corpus linguistics: A mutualistic entente* (pp. 191–204). Berlin/New York: Mouton de Gruyter.

Lorenzo-Dus, N. (2001). Compliment responses among British and Spanish university students. A contrastive study. *Journal of Pragmatics, 33*, 107–127.

Lorenzo-Dus, N., & Meara, P. (2004). Role-plays and the assessment of oral proficiency in Spanish. In R. Márquez-Reiter & M. E. Palencia (Eds.), *Current trends in the pragmatics of Spanish* (pp. 79–98). Amsterdam/Philadelphia: John Benjamins.

Lozano, C., & Mendikoetxea, A. (2011). CEDEL2 (Corpus Escrito del Español como L2): A large-scale corpus for L2 Spanish acquisition research. In M. L. Carrió Pastor & M. A. Candel Mora (Eds.), *Proceedings of CILC 2011. Universidad Politécnica de Valencia (Spain)*. Valencia: Universidad Politécnica de Valencia.

Martín Zorraquino, M. A., & Portolés, M. A. (1999). Los marcadores del discurso. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (Vol. 2, pp. 4051–4214). Madrid: Espasa-Calpe.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies. An advanced resource book*. London/New York: Routledge Applied Linguistics.

Miracle, W. C. (1991). *Discourse markers in Mandarin Chinese*. PhD thesis, Ohio State University, Columbus. Available at: https://etd.ohiolink.edu/ap:10:0::NO:10:P10_ACCESSION_NUM:osu1232114951. Accessed 27 Aug 2013.

Mitchell, R., Domínguez, L., Arche, M. J., Myles, F., & Marsden, E. (2008). SPLLOC: A new database for Spanish second language acquisition research. In L. Roberts, F. Myles, & A. David (Eds.), *EuroSLA yearbook* (Vol. 8, pp. 287–304). Amsterdam/Philadelphia: John Benjamins.

Moreno, A., & Guirao, J. M. (2006). Morpho-syntactic tagging of the Spanish C-ORAL-ROM corpus: Methodology, tools and evaluation. In Y. Kawaguchi, S. Zaima, & T. Takagaki (Eds.), *Spoken language corpus and linguistic informatics* (pp. 199–218). Amsterdam/Philadelphia: John Benjamins.

Mukherjee, J. (2009). The grammar of conversation in advanced spoken learner English: Learner corpus data and language-pedagogical implications. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 203–230). Amsterdam/Philadelphia: John Benjamins.

Müller, S. (2005). *Discourse markers in native and non-native English discourse*. Amsterdam/Philadelphia: John Benjamins.

Nogueira da Silva, A. (2011). *Enseñanza de los marcadores del discurso para aprendices brasileños de E/LE: análisis de manuales de E/LE y de narraciones orales de aprendices brasileños*. PhD thesis, Universidad de Salamanca, Salamanca. Available at: http://hdl.handle.net/10366/110547

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom. Language use and language teaching*. Cambridge: Cambridge University Press.

Polat, B. (2011). Investigating acquisition of discourse markers through a developmental learner corpus. *Journal of Pragmatics, 43*(15), 3745–3756.

Pons Bordería, S. (1998). *Conexión y conectores. Estudio de su relación en el registro informal del la lengua*. Valencia: Universidad de Valencia.

Pons Bordería, S. (2006). A functional approach to the study of discourse markers. In K. Fischer (Ed.), *Approaches to discourse particles* (pp. 77–99). Oxford: Elsevier.

Porroche Ballesteros, M. (1998). Sobre algunos usos de *que*, *si* y *es que* como marcadores discursivos. In A. Mª, M. Zorraquino, & E. Montolío Durán (Eds.), *Los marcadores del discurso. Teoría y práctica* (pp. 229–242). Madrid: Arco/Libros.

Ramírez Verdugo, D., & Romero-Trillo, J. (2005). The pragmatic function of intonation in L2 discourse: English tag questions used by Spanish speakers. *Intercultural Pragmatics, 2*(2), 151–168.

Romero-Trillo, J. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics, 34*(6), 769–784.

Romero-Trillo, J. (2004). Subjective and objective grounding in discourse markers: A cross-linguistic corpus-driven approach. In K. Aijmer (Ed.), *Dialogue analysis VIII: Understanding and misunderstanding in dialogue: Selected papers from the 8th IADA conference, Göteborg 2001* (pp. 185–197). Tübingen: Niemeyer.

Romero-Trillo, J. (Ed.). (2008). *Pragmatics and corpus linguistics. A mutualistic entente* (Mouton Series in Pragmatics, Vol. 2). Berlin: Mouton de Gruyter.

Romero-Trillo, J. (Ed.). (2013). *Yearbook of corpus linguistics and pragmatics 2013. New domains and methodologies*. New York: Springer.

Rose, K., & Kasper, G. (2001). *Pragmatics in language teaching*. Cambridge: Cambridge University Press.

Santos Río, L. (2003). *Diccionario de partículas*. Salamanca: Luso-Española de ediciones.

Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.

Schmer, D. (2012). Un análisis de nada como marcador del discurso en el español coloquial de Buenos Aires. *Revista del Instituto de Investigaciones Lingüísticas y Literarias Hispanoamericana (RILL), 17*(1/2), 218–227.

Sessarego, C. (2009). Pragmatic language instruction and beginner learners of Spanish: A discourse approach to pragmalinguistics. *Estudios de Lingüística Aplicada, 49*, 97–120.

Simon-Vandenbergen, A.-M., & Willems, D. (2011). Cross-linguistic data as evidence for grammaticalization. In L. Degand, & A.-M. Simon-Vandenbergen (Eds.), *Grammaticalization,*

pragmaticalization and/or (inter)subjectification: Methodological issues for the study of discourse markers. Thematic Issue of *Linguistics*, *49*(2): 333–364. Berlin: De Gruyter Mouton.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell's.

Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics, 4*(2), 91–112.

Walsh, S. (2013). Corpus linguistics and conversation analysis at the interface: Theoretical perspectives, practical outcomes. In J. Romero-Trillo (Ed.), *Yearbook of corpus linguistics and pragmatics 2013. New domains and methodologies* (pp. 37–51). New York: Springer.

Wei, M. (2011). Investigating the oral proficiency of English learners in China: A comparative study of the use of pragmatic markers. *Journal of Pragmatics, 43*(14), 3455–3472.

Yao, J. (2012). Clasificación de los marcadores del discurso en chino mandarín. In F. J. Rodríguez Muñoz (Ed.), *Los marcadores del discurso en español y en chino mandarín. Philologica Urcitana. Revista Semestral de Iniciación a la Investigación en Filología* (Vol. 7, pp. 5–20). Almería: Universidad de Almería.

Yoshimi, D. R. (2001). Explicit instruction and JFL learners. Use of interactional discourse markers. In K. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 223–244). Cambridge: Cambridge University Press.

Zhao, H. (2013). A study on the pragmatic fossilization of discourse markers among Chinese English learners. *Journal of Language Teaching and Research, 4*(4), 707–714.

"Hope This Helps!" An Analysis of Expressive Speech Acts in Online Task-Oriented Interaction by University Students

Marta Carretero, Carmen Maíz-Arévalo, and M. Ángeles Martínez

Abstract This paper explores the presence of expressive speech acts in a corpus of e-forum history logs derived from the online collaborative writing activity of three groups of undergraduate and postgraduate students in a tertiary education setting. The macro category of Expressives has been less frequently studied than others such as Directives or Commissives, and even nowadays its in-depth study tends to concentrate on specific subtypes such as Compliments. In computer mediated exchanges, the implicit disembodiment must ensure an outstanding role for expressive uses of language, since non-verbal means are not available as in face-to-face conversation. The study includes a qualitative and quantitative analysis which covers the similarities and differences found across the subcorpora corresponding to each of the three groups of students involved, in terms of subtypes of Expressives and their linguistic realisations. The results suggest that Expressives play a crucial role as rapport building devices in the online interaction, smoothing and complementing transactional language. The analysis also suggests that the variables of linguistic proficiency, group size, age, multiculturality, and method of assessment may have a bearing on the form and use of Expressives in online written interaction in blended learning environments.

Keywords Expressive speech acts • Online collaborative learning • E-forums • Tertiary level • Emoticons

M. Carretero (✉) • C. Maíz-Arévalo • M.Á. Martínez
Departamento de Filología Inglesa I, Facultad de Filología – Edificio A, Universidad Complutense – Ciudad Universitaria, E-28040, Madrid, Spain
e-mail: mcarrete@filol.ucm.es

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2,
DOI 10.1007/978-3-319-06007-1_12, © Springer International Publishing Switzerland 2014

1 Introduction

The analysis of expressive speech acts presented in this article is based on the log history of an institutional e-forum used for several collaborative writing activities with tertiary level students at the Complutense University of Madrid, Spain. Collaborative writing can be defined as “the joint production of a text by two or more writers” (Storch 2011: 275) and its defining characteristic is the co-authorship of the text produced. Online collaboration in text production, on the other hand, differs from face-to-face interaction in two key respects. In the first place, as opposed to face-to-face exchanges, online communication may be of an asynchronous nature, as in the case of blogs, wikis, and e-forums. On the other, the interaction itself, as well as the evolving versions of the emerging product, are often recorded in detail by the social computing medium itself, in the form of a log history. This log is actually a complete transcription of the linguistic production of participants, and is therefore an invaluable source of information on the language and pragmatics of online collaboration.

1.1 Characteristics of Online Collaborative Writing

The use of online collaborative writing for pedagogical purposes is inextricably linked to the boost of collaborative learning in the 1990s (Dillenbourg et al. 1996, Dillenbourg 1999), and is in turn rooted in the socio-cultural and interactionist theories of learning derived from the works of Piaget (1928) and Vygotsky (1978). The advantages of collaborative learning have been frequently mentioned in the literature. As Neumann and Hood (2009: 383) point out, “Collaborative learning has been associated with higher achievement, higher motivation, positive student-student relationships, and more positive attitudes towards the discipline of study.” Other frequently observed beneficial effects are increased student involvement with the subject matter, and better critical thinking and problem-solving skills (Cole 2009: 143). In the case of online collaborative writing, an additional advantage is an increase in learners’ individual autonomy, as the technology used in this kind of collaboration promotes learner control over the learning process and forces a shift in the roles of teachers and students (Kessler et al. 2012: 92), empowering learners and placing them at the centre of the learning process (Blake 2011: 25; Leeming and Danino 2012: 54). As a result, participants in the collaboration may eventually be involved in a variety of non-exclusive, and often overlapping roles that include writer, consultant, editor, reviewer, leader, and facilitator (Nöel and Robert 2004: 66).

The varied role-relationships that these configurations entail must no doubt pose great social demands on participants, as their linguistic repertoires in terms of pragmatic knowledge and skills should include a rich array of linguistic devices needed to successfully cope with the social aspects of online collaboration. As early as 1999, in his pioneering studies on collaborative learning, Pierre Dillenbourg (1999) already noted that scientific approaches to collaborative learning, probably

influenced by the role of computer scientists in its development, tended to focus on the cognitive aspects and benefits of collaboration, but "deliberately neglected the socio-affective aspects of collaborative learning" (Dillenbourg 1999: 14). It may be argued that the physical detachment imposed by the technology would cool down the social pressure of a face-to-face learning setting, but research in computer mediated communication (Herring et al. 2013), however, suggests that it is precisely this distance that has to be linguistically cut short, as language alone has to make up for the lack of non-verbal information in the form of gestures and intonation. This is why the use of expressive speech acts can be predicted to be of paramount importance in the socially-driven parts of the linguistic production of participants in online collaborative writing activities. By "zooming in" in this linguistic production using the permanent record allowed by the e-forum log, it should be possible to focus on the form and function of these speech acts in an online collaborative setting, thus throwing light on some of the socio-affective and pragmatic aspects of online collaboration.

1.2 Aim of the Study

The aim of this study is, thus, to explore the use of expressive speech acts in online written collaboration as medium-determined strategies fulfilling the communicative function of building and maintaining rapport, simultaneously smoothing the interaction and driving it to satisfactory task completion and social harmony. The three groups of students participating in the experience were taking courses on English Studies at a tertiary level. The groups were subdivided into several three- to four-member smaller groups for the online collaborative writing activities, and all of them used the institutional university e-forum as the medium of interaction (see Sect. 3 for more details). But, apart from these similarities, their internal composition and situational features differed in several important respects, and this was, precisely, what prompted the following research questions:

(a) Is the presence of expressive speech acts homogeneous across the three subcorpora, in terms of overall frequency and of the presence of sub-types such as Thankings, Apologies, Compliments, and so forth?
(b) If this is not the case, how can the existing differences be connected to relevant contextual features with a bearing on participants' linguistic choices?
(c) Does the fact that the online collaboration took place in a blended learning environment, combining face-to-face and e-forum interaction, in any way affect the use and selection of Expressives by participants?

The paper will first specify the concept, scope and types of expressive speech acts adopted, followed by a description of the methodology and contextual features of the research, to finally focus on the quantitative and qualitative analysis of the 79,699-word long corpus containing the online interaction of the three macro-groups across the different e-forum tasks they performed along a 2-week time period. The study shows that there are remarkable differences in the use of

expressive speech acts by the three participating groups, and that these differences can be connected to contextual features, particularly to a set of five variables – proficiency level, age, cultural background, group size, and assessment method – which seem to have a statistically sound bearing on the results as demonstrated by an ANOVA test (see Sect. 4.1).

2 Expressive Speech Acts: Concept, Types and Scope

2.1 General Considerations

Expressives are one of the basic speech act types proposed in Searle's (1976) seminal classification, together with Representatives, Directives, Commissives and Declaratives. In the literature, the scope of Expressives has been a matter of debate. Not in vain did Austin (1975) consider his category of 'behavitives', a precursor of Expressives, as "miscellaneous" and "troublesome" (1975: 152). Austin described behavitives as having to do with attitudes and social behaviour, and gave examples such as "apologizing, congratulating, commending, condoling, cursing, and challenging" (1975: 152). Searle, in an attempt to provide a more accurate classification of speech acts, characterizes Expressives basically by means of two features. The first is their illocutionary point, described as "to express the psychological state specified in the sincerity condition about a state of affairs specified in the propositional content" (1976: 12). The second is the lack of direction of fit, in the sense that a speaker uttering an Expressive is not trying to match the world with the words nor the words with the world; instead, the proposition is presupposed. Searle gives Apologize, Congratulate and Thank as examples of Expressives, and states that the fact that the truth of the proposition is presupposed is reflected in the syntax of English: significantly, the performative verbs corresponding to the speech acts mentioned above cannot take *that*-clauses, but require other constructions. Searle (1976: 12) demonstrates this with examples (1–3) (the numbering and pairing is the authors'):

(1)	(a) *I apologize that I stepped on your toe.[1]
	(b) I apologize for stepping on your toe.
(2)	(a) *I congratulate you that you won the race.
	(b) I congratulate you on winning the race.
(3)	(a) *I thank you that you paid me the money.
	(b) I thank you for paying me the money.

[1] The examples with no indications have been constructed by the authors or adopted from a different source from the corpus (indicated in all the cases). The examples cited from the corpus are followed by an indication between brackets of the subcorpus to which they belong.

Searle limits his examples of Expressives to these performatives. It may well be argued that the absence of direction of fit is due to the fact that the examples of Expressives chosen are realized by metalinguistic performative verbs, which are always true and also felicitous and successful (Thomas 1995: 36). If these meta-linguistic performatives are replaced with other expressions that communicate the same speech act, the resulting sentence has truth conditions, and hence the direction of fit is from words to world. For example, if (1b) is replaced with (4), which is also an Apology, the speaker presupposes that s/he stepped on the addressee's toe, but also claims for the truth that s/he is sorry about that. Moreover, Searle's vision of Expressives seems quite restricted: he does not mention cases in which the speaker's psychological state is communicated by different means from lexical items expressions, such as syntax or intonation; for instance, (5–6) would be intuitively classified as Expressives:

(4)	I am sorry that I stepped on your toe.
(5)	What a nice coat you're wearing!
(6)	Silly me!!

We believe that the root of the problems posed by the definition and scope of Expressives lie in their association with psychological states. Psychological states are not exclusive to Expressives, but pervasive in all kinds of speech acts. On the one hand, it may be argued that all speech acts express psychological states. As Verschueren (1999: 132) states, the relevance of the very class of Expressives can be denied according to Searle's own criteria for classification. If Expressives are defined as those speech acts that express psychological states, then they include all Assertives, Directives and Commissives, since these kinds of speech acts express a belief, a wish and an intention, respectively. Likewise, psychological states are also pervasive in Weigand's dialogic speech act taxonomy, entirely built on the basic mental states of belief and desire (Weigand 2010: 83). On the other hand, emphatic expression of a psychological state is not incompatible with any of the other kinds of speech acts: for example, any speech act type may be uttered with an intonation that unmistakably communicates joy, sorrow, anger or surprise, as would be the case of the Statement in (7), the Question in (8) or the Directive in (9) uttered angrily:

(7)	Jim has left me for my best friend.
(8)	Where did you leave my pen drive?
(9)	Stop treading on my toes.

2.2 Classifying Expressive Speech Acts

The present study requires a fine-grained classification of Expressives into subtypes that calls for a brief preliminary discussion of the theoretical groundings behind the analysts' decisions. In this paper, Expressives will be considered to concern the basic

mental state of desire rather than that of belief. Therefore, those kinds of speech acts defined in terms of belief, such as Assertives, Constatives, Predictives or Suppositives, among many others, will not be considered as Expressives. However, the categories of belief and desire frequently cut across each other, and, in this respect, Expressives may present intra-category differences as far as belief is concerned.

Among the speech acts that focus on the mental state of desire, we will not consider as Expressives per se those types in which the fit from world to word is the primary illocutionary point, such as Directives or Commissives. Evaluations, on the other hand, may be considered as Expressives in the sense that they communicate positive or negative perceptions of entities or states of affairs. But, although evaluation tends to be associated with reason and emotion with sentiment, the difference between them is far from clear: rather than being separate categories, they seem to make up a continuum. For example, in uttering (10), a speaker may well communicate both rational evaluation and emotion. However, for our classification of Expressives, we believe that the emotional is to be given priority. In this respect, the approach to evaluative language proposed by the Appraisal framework (Martin and White 2005; White 2003) provides a useful distinction between Affect, Judgement and Appreciation. Affect is defined as "concerned with registering positive and negative feelings: do we feel happy or sad, confident or anxious, interested or bored?" (Martin and White 2005: 42). (11–13) are sentences that contain expressions of Affect. In contrast, Judgement covers moral evaluations of character (14) and Appreciation captures practical or aesthetic qualities of objects and natural phenomena (15). Affect pertains more to the emotional, while Judgement and Appreciation are more based on reason. According to this difference, explicit expressions of Affect concerning the writer's feelings, such as (11–13), will be considered as Expressives, while those of Judgement and Appreciation will not. Among these Expressives, a distinction will be made between 'Liking', which expresses positive affect, as in (11–12), and 'Concern', which expresses negative affect, as in (13):

(10) The colours of this painting are really beautiful.
(11) I'm very happy about you.
(12) Fortunately, I managed to finish the paper in time.
(13) I'm worried about the results of the analysis.
(14) Your boss is a reliable person.
(15) This novel is interesting indeed.

Our classification of Expressives will also include Weigand's 'desideratives' (2010: 171, 201–203), which make a modal claim to truth: in Givón's terms (1982: 24; 1984: 252–256), the proposition is irrealis. Desideratives claim that the truth of the proposition should (or should not) be the case, but at the speech time they are not true (16) or their truth is only a possibility (17–18). The label 'Wish' will be used for these acts.

(16) I wish I had worked harder this year.
(17) Hopefully, she will work harder from now on.
(18) I fear that the results will not be terribly good.

Liking, Concern and Wish might be labelled as ‘self-centred’, since they pertain to the speaker/writer’s feelings. In contrast, other Expressives are ‘other-centred’ in that they focus on the addressee. These include the Expressives originally specified in Searle (1976), such as Apologies, Thankings and Compliments. Weigand (2010: 179) considers these as declaratives, since they refer to routines of behaviour in which authentic feeling is not always present and their main purpose is the creation of social relationships by the use of politeness conventions. It is true that in many cases these speech acts have a high degree of conventionalization which often bleaches authentic feeling, as stated by Austin himself (1975: 80–81) and corroborated by other authors (Bach and Harnish 1979: 51). However, we still believe that these acts have a strong element of concern towards the addressee, since they are clearly aimed at keeping a good rapport with him/her, and will consequently consider them as Expressives. These acts may be expressed by performative verbs but also by other means, as in (19–20):

(19) I am grateful for your help.
(20) Congratulations!!

Apologies are sometimes followed by Reassuring speech acts, by which the speaker/writer aims at comforting the addressee by diminishing his/her feeling of guilt. Examples of conventionalized Reassuring expressive speech acts are *Never mind*, *It’s all right* or *Don’t worry*. Another kind of other-centred Expressives are Reproaches (Weigand 2010: 203–205), which may be seen as the negative counterpart of Compliments, since they express a negative evaluation of the addressee.

Finally, our scope of Expressives also includes speech acts of different kinds from those mentioned above which focus on the speaker/writer’s emotional involvement by other means. The first is the exclamative grammatical construction (21–22). Weigand (2010: 166) cites this construction as an example of her ‘emotive’ speech acts, which she defines as “focus[ing] on the speaker’s *emotional involvement*” (her italics). It must be noted that, with this construction, the truth of the proposition is presupposed.

(21) What an extraordinary picture this is!
(22) How extraordinary this picture is!

Weigand (2010: 166–167) states that, besides the exclamative sentence type and intonation, interjections such as *oh*, particles and routine phrases are typical devices of emotives. Along these lines, we consider that certain orthographic or typographical devices, such as exclamation marks, interjections, typographical repetition of a letter, capitalization, change of colour or underlining, and also the use of emoticons (non-linguistic signs) focus on the writer’s emotional involvement (Yus 2011), and hence utterances containing any of these marks will be considered as Expressives, independently of the kind of speech act that they realize. Repetition of the question mark sign will also be considered as a feature of Expressives: this repetition puts emphasis on doubt (a subtype of belief), but also indicates emotional involvement

about this doubt, which motivates our classification of these speech acts within Concern. This feeling is well characterized by Givón's (1990: 529) term 'epistemic anxiety', as in example (23), taken from the data:

(23) What is the PCI of: look at the papers on his desk??? (Pr[2])

Other kinds of speech acts are often reinforced in the ways expressed above. Occurrences were found of Greetings and Farewells, which turned out to be an important category of other-centred Expressives that create or maintain social relationships and, in so doing, express concern towards the addressee. Both have been grouped under Greetings. There were also examples of typographically marked Assertions, Commissives and Directives. Due to its importance in the three subcorpora, the category of Agreement will be split from the rest of the statements and conferred the status of another subtype. Examples were also found with emphatic *do*, as well as accumulation of intensifiers and evaluative expressions. It may well be thought that these devices enhance emotional involvement, and consequently the acts containing them have been considered as Expressives too. This is the case of (24), which contains two emotive speech acts of Agreement:

(24) I *totally* agree with you. I think this is a *perfect* idea. (SL) (italics added by authors).

Table 1 summarizes the classification of Expressives used in this paper, based on the considerations above, and illustrated with examples from the three subcorpora.

3 Methodology

3.1 Participants

The corpus on which this study is based consists in the online written interaction of 83 undergraduate and postgraduate students enrolled in three different courses within the English Linguistics programme at the Complutense University of Madrid, Spain. The largest group consisted of 64 undergraduate students taking a course in the optional subject *Discourse and Text* (D&T henceforth), while 9 were also undergraduate students from an evening group following an obligatory course on *Pragmatics* (Pr henceforth). Finally, the remaining 10 were post-graduate students doing a Master's seminar called *Seminar on English Linguistics* (SL henceforth). The gross number of undergraduates in morning groups is clearly superior to that of undergraduates in evening obligatory subjects, and also to that of

[2] Each of the subcorpora has been named according to the subject the students were doing. The acronyms stand as follows: Pr stands for Pragmatics (evening undergraduate group), D&T for Discourse and Text (morning undergraduate group) and SL for Seminar of Linguistics (master's group). See the Methodology section for detailed information.

Table 1 Classification of Expressives in this study

Speech act type	Example
Liking	I really like the classification. (SL)
Concern	I cannot recognize PCIs nor GCIs. . . It is difficult to see them. . . the easiest are the presuppositions xD (Pr)
Wish	I hope I’ve copy-pasted everything in the right place. (SL)
	I wanted to answer to the last part of question two and question three but I really cannot think any longer. (Pr)
Apology	Excuse me for my delay. (D&T)
	I’m sorry girls I had to work the whole morning, I’m going to look at the document immediately! :-D (SL)
Reassuring	Don’t worry because everything is finished and sent (D&T)
Thankings	Thanks a lot! (D&T)
Compliment	I think all the comments you´ve added are interesting. (SL)
Reproach	I feel like I’m having pretty much of a monologue here. . . (D&T)
Greeting/ farewell	Hello everybody! (D&T)
	See you in class! :-) (SL)
Emotive assertion	I have finished my part! (D&T)
Emotive directive	Suggestions would be very welcome!! (Pr)
Emotive commissive	I’m going to try to post my ideas tomorrow! (D&T)
Emotive agreement	I agree with everything you’ve said:D (SL)

postgraduates, a fact that could not be controlled for the present research. As for participants’ age, students in the morning group are in their early twenties on average, with slightly older students in the evening group, who are in their late twenties. Finally, in the case of the master’s students, age ranges from mid-twenties to early forties. This also establishes an age variable to be considered when analysing the results. On the whole, morning groups usually consist of students who do not work elsewhere as opposed to evening and master’s students, who often combine work and studies and are therefore more used to taking responsibility and maximising their efficiency.

As for their level of English, it was fairly advanced in general terms. However, the morning group can be said to have a slightly lower level (despite some exceptions), in terms of both linguistic ability and academic background, being also a mixed-ability group where students from second to fourth academic years co-existed. Their level ranged from B2 to C1 according to the *Common European Framework of Reference* (2001). With regard to the evening group of undergraduates, this was a specially “gifted” group formed by extremely hard working and devoted students, also with a level ranging from B2 to C1. In contrast to the undergraduate students, the master’s participants were highly proficient in English, some of them being teachers of English as a foreign language in their countries of origin, and their overall level ranged between C1 and C2.

Besides, it is also necessary to mention that the first two groups consisted largely of students whose mother tongue was Spanish (despite some minor exceptions) whilst the third group was integrated by international students of different mother tongues (e.g. Russian, Korean, Arabic, Polish and Spanish). Once again, this difference was particularly interesting inasmuch as it permitted an examination of the effect of intercultural differences.

In order to avoid unnaturally biased exchanges, the participants were not informed a priori of their participation in this research project; once the experiment was over, they were dutifully informed and asked for their written consent, even if pseudonyms were used and all the data were kept anonymous to protect their identities.

3.2 Procedure

Each group of participants was asked to carry out one or two collective assignments, and were specifically required not to do these collaborative exercises in the traditional face-to-face way but online, by means of the e-forum tool provided by the Moodle platform used by the university's Virtual Campus. As for the assignments, they were distributed as follows. The first, and also the most numerous group, did just one assignment. The students were requested to examine a text according to the parameters of Register Theory; more specifically, a one-page article from the magazine *Time* had to be explored along the register parameters of Field and Tenor, using the linguistic tools seen in the class. This analysis led to a final report answering a set of questions provided by the teacher.

As opposed to the first group, the other two (significantly less numerous) were required to do two assignments each so that a more comparable amount of data was collected. With regard to the undergraduate students of Pragmatics, their first activity consisted in the detection in written texts of presuppositions and conversational implicatures, accordingly justifying their responses. Their second project consisted in the identification and further analysis of themes and local topics in three naturally-occurring written texts,[3] accompanied by answers to some questions about the stylistic effects of the different kinds of themes and the distribution of the local topics.

Finally, the master's students also carried out a first task which consisted in analysing a series of multimodal texts belonging to children's illustrated fairy tales and their impact on gender construction. After the analysis, students were to upload a final report. The second activity was intended to assess their pragmatic awareness. They were requested to examine a small corpus of naturally occurring data and write a report on their analysis of intercultural pragmatic differences and pragmatic failure.

[3] The selected texts were the beginnings of "The sisters", "An encounter" and "Eveline", three short stories included in James Joyce's Dubliners (1914).

Table 2 Variables at play

Variable	Morning group	Evening group	Master's group
Group size	64 students	9 students	10 students
Age	Early 20s	Late 20s	Mid- 20s/early 40s
Linguistic proficiency	B2-C1	B2-C1	C1-C2
Cultural homogeneity	Monocultural	Monocultural	Multicultural
Assessment method	E-forum assessed	E-forum participation	E-forum not assessed
	Both individual and collective mark	Individual mark	Collective mark

The assessment method of these tasks was carried out in different ways. The morning group was evaluated along the double axis of individual online participation and final assignment assessment. Students received two grades for this assignment: an individual grade for each student and a common grade for all group members. The students of the evening group received only a mark, mainly based on individual participation although the final uploaded report was also considered. As for the master's students, they received only a collective mark based on their final uploaded report. E-forum participation was not considered in this case.

Summing up, five different variables seem to have a possible bearing on the results, as illustrated by Table 2.

The three subgroups followed the general policy of limiting the number of participants per e-forum to a maximum of four. Thus, the first group (64 students) resulted in an average of four members per e-forum (16 e-forums); the second and third groups encompassed an average of three members per e-forum. Likewise, the groups were, in all cases, randomly created, either by the Virtual Campus itself or by using alphabetical order. A third common feature was the amount of time allotted for the realization of the activities, limited to a maximum of 2 weeks before the actual uploading of the students' final reports.

3.3 Description of the Corpus

The data used in the study consists in a 79,699-word long corpus made up of three subcorpora, each of them containing the e-forum written interaction of the three groups of students described above. Subcorpus A (D&T) consists of 40,226 words. Subcorpus B (Pr) of 14,119, subdivided into 7,736 of the first e-forum and 6,383 of the second. Finally, Subcorpus C (SL) comprises a total of 25,354 words, also subdivided into 15,598 for the first e-forum and 9,749 for the second.

As for the unit of analysis, the need to unify criteria led us to choose not the discursive paragraph, more open to a qualitative analysis, but the sentence, more

suited to our quantitative purposes in the present paper, with due attention to the conversational implicatures derived from context and co-text. But working at sentence level does not rule out the fact that speech acts are derived from the meaning of sentences as utterances in specific contexts of situation. It is precisely this attention to contextual features with a bearing on speech act disambiguation that determines the need to have access to non-immediate co-text in order to assign a specific pragmatic value to a given sentence. This need drastically limits, in our view, the effectiveness of computational tools for this particular study. The retrieval of the occurrences of given words or expressions by means of a concordancer would be ideal for quantitative analyses focusing on recurrent formal realizations of speech act types, such as "sorry" for Apology or "Thank you" for Thanking. However, this procedure brings to mind Weigand's (2010) distinction between empiricist and functionalist approaches to speech act taxonomising:

> Whereas the *empiricists* maintain that the only access to speech act types is by means of expressions – speech act verbs, sentence types and categories of modality -, the *functionalists* stress the point that speech acts are intensionally functional concepts. [...] Verbal expressions and speech acts are different categories. Verbal expressions are an integrated component of the utterance. Speech acts are defined at the functional level by their purpose. (Weigand 2010: 132)

Given the analysts' privileged access to the context of situation in this study, we believe that it is worth carrying out a functionalist rather than an empirical analysis, which would have the disadvantage of limiting the search to form, thus excluding function (Adolphs 2008: 9). In our case, the data collectors and analysts were also a part of the communicative situation, as activity designers and interaction participants themselves. Therefore, the quantitative analysis is minimally constrained by the shortcomings often signalled in the literature that considers the pros and cons of the use of corpora for the analysis of speech acts (Rühlemann 2010: 288–291; Weigand 2010: 27–28, 40), such as incomplete knowledge of the situation in which the discourse is produced and lack of access to surrounding non-linguistic features or non-textual features such as font size and type, line breaks, or the use of emoticons. For these reasons, the classification of the Expressives in our corpus was handled manually. In fact, the kind of intention likely to have motivated the use of a certain utterance in a specific context of situation cannot always be straightforwardly associated to the kind of conventionalized forms easily spotted by mechanical instruments. John Sinclair himself, in his seminal *Corpus, Concordance, Collocation,* notes that there are "instances which do not easily detach from their contexts, or which require a very extensive stretch of text to avoid distortion" (Sinclair 1991: 5). The manual analysis carried out in this study uncovered many such cases. For example, (25) gives no clear formal clues about the kind of Expressive it is: it could well be a Reproach or a Compliment, for instance, but the previous linguistic context indicates that it is an Apology about a response that the writer had placed under a wrong question number in a previous turn:

(25) It’s actually question 4!!! (Pr)

It must also be noted that, in the course of the analysis, the categorization of certain speech acts was far from easy, even when access to the complete exchange was granted. Difficulties arose not only when deciding among subtypes of Expressives, but also when deciding whether certain acts were expressive or not. Consider examples 26, 27, 28, and 29. 26 is a hybrid case in which Apology, Concern, and Reproach seem to merge. The three of them convey uneasiness on the part of the speaker, the difference lying on the fact that Apologies express that the speaker feels guilty towards the others, Reproaches express that the speaker feels that the others are guilty towards her or him, and Concerns orient the uneasiness towards the speaker’s inner mood:

(26) Sorry but I have been waiting during 2 weeks so that someone could answer me and nobody did (D&T).

Liking and Compliment are also frequently difficult to disambiguate, as in 27, while Compliments frequently merge with mere Agreement, as in 28:

(27) I like your suggestion (SL).
(28) OK, I think your analysis is very complete and very good (SL).

Finally, speech acts of Concern were also often difficult to distinguish from non-expressive informative acts of knowledge, as in 29:

(29) I am really stuck with this activity (D&T).

Bearing these ambiguities in mind, in dubious, hybrid cases, the decision was based on the apparent primary function of a specific act, but this was not always without problems, as secondary though not minor meanings had to be sacrificed for the sake of clarity in the quantification. Hence, once the data were manually searched, the utterances that fulfilled the criteria for Expressives presented in Sect. 2 were classified in an Excel table, so as to unify the three subcorpora. All in all, a total of 842 expressive speech acts were analysed, subdivided as follows: 476 (D&T), 133 (Pr) and 233 (SL).

In order to triangulate the data, each of the three researchers carried out a preliminary individual analysis of her own dataset. The findings were then compared and discussed until the criteria of analysis under every subtype were ratified by the three members of the research team. The three resulting Excel tables were revised by the three researchers. The most significant results found after the analysis will be discussed in the following Section.

Table 3 Cross-comparative view of results

Speech acts	Pr subcorpus %	SL subcorpus%	D&T subcorpus%
Apology	25.56 (N = 34)	10.72 (N = 25)	10.90 (N = 52)
Compliment	16.54 (N = 22)	21.00 (N = 49)	14.89 (N = 71)
Greeting	9.02 (N = 12)	13.73 (N = 32)	16.14 (N = 77)
Wish	6.77 (N = 9)	3.43 (N = 8)	17.20 (N = 82)
Thanking	6.77(N = 9)	18.88 (N = 44)	19.91 (N = 95)
Liking	0.00 (N = 0)	4.29 (N = 10)	0.42 (N = 2)
Concern	10.53 (N = 14)	1.71 (N = 4)	2.93 (N = 14)
Reproach	4.51 (N = 6)	0.85 (N = 2)	5.66 (N = 27)
Directive	7.52 (N = 10)	13.30 (N = 31)	4.40 (N = 21)
Agreement	4.51 (N = 6)	3.00 (N = 7)	3.14 (N = 15)
Assertion	3.01 (N = 4)	6.86 (N = 16)	1.44 (N = 7)
Commissive	2.26 (N = 3)	1.71 (N = 4)	2.10 (N = 10)
Reassuring	3.01 (N = 4)	0.42 (N =1)	0.63 (N = 3)
Total	100 (N = 133)	100 (N = 233)	100 (N = 476)

4 Data Analysis and Discussion of Findings

4.1 Subtypes of Expressives

By way of introduction, Table 3 offers a panoramic view of the differences with which participants in each of the groups have employed the Expressives under analysis. Each subtype is accompanied by the number of tokens (N), together with the corresponding percentage when compared with the total number of Expressives in each of the three subcorpora.

Given that the data resulted from three independent subcorpora, the use of an ANOVA F-test was considered the most appropriate way to verify the significance of the results. The ANOVA test produces an F-statistic, the ratio of the variance calculated among the means to the variance within the subcorpora, where the variance between the group means should be lower than the variance of the subcorpora, following the central limit theorem. In this case, the analysis yielded an F-value of 4.78, clearly higher than the critical value ($F_{crit} = 3.25$ at $\alpha = 0.05$). Since $F = 5.36 > 3.25$, the results are significant at the 5 % significance level. Tables 4, 5 and 6 sum up the data entry, the total standard deviation and the deviation across groups, respectively.

Zooming in on the results shows that the four most frequently performed Expressives, in descending order, in each of the groups are those specified in Table 7. As can be observed, similarities and differences are found between the three groups, with Greeting and Compliment being the Expressives used with relatively the same frequency, while Apology, Concern, Directive and Wish appear only in one of the subcorpora. For the sake of clarity, the rest of the section has been divided into two major subsections, respectively focusing on the similarities and

Table 4 Data entry

Speech acts	Pr	SL	D&T	TOTAL
Apology	34	25	52	111
Compliment	22	49	71	142
Greeting	12	32	77	121
Wish	9	8	82	99
Thanking	9	44	95	148
Liking	0	10	2	12
Concern	14	4	14	32
Reproach	6	2	27	35
Directive	10	31	21	62
Agreement	6	7	15	28
Assertion	4	16	7	27
Commissive	3	4	10	17
Reassuring	4	1	3	8
Total	133	233	476	842
Mean	10.23	17.92	36.62	
Total mean			21.59	

Table 5 Total standard deviation

	$(X_{1i}-AVG)^2$	$(X_{2i}-AVG)^2$	$(X_{3i}-AVG)^2$
Apology	154.01	11.63	924.78
Compliment	0.17	751.32	2,441.37
Greeting	91.96	108.37	3,070.30
Wish	158.50	184.68	3,649.40
Thanking	158.50	502.22	5,389.07
Liking	466.12	134.32	383.76
Concern	57.60	309.40	57.60
Reproach	243.04	383.76	29.27
Directive	134.32	88.55	0.35
Agreement	243.04	212.86	43.42
Assertion	309.40	31.25	212.86
Commissive	345.58	309.40	134.32
Reassuring	309.40	423.94	345.58
Total	2,671.65	3,451.70	16,682.09
Std. Dev.			22,805.44

differences found in the three subcorpora when comparing the frequency of use of the subtypes of Expressives.

4.2 Similarities Across the Subcorpora

The analysis of the Expressives uncovers certain similarities, from the quantitative and qualitative points of view. The most salient feature is, perhaps, the predominance of other-oriented over self-oriented speech acts. Compliment and Greeting rank among the four most frequent subtypes in the three subcorpora. Thanking also

Table 6 Deviation across groups

	$(X_{1i}-AVG_1)^2$	$(X_{2i}-AVG_2)^2$	$(X_{3i}-AVG_3)^2$	
Apology	564.98	50.08	236.69	
Compliment	138.51	965.78	1,182.30	
Greeting	3.13	198.16	1,630.92	
Wish	1.51	98.47	2,059.76	
Thanking	1.51	680.01	3,408.76	
Liking	104.67	62.78	1,198.22	
Concern	14.21	193.85	511.46	
Reproach	17.90	253.54	92.46	
Directive	0.05	171.01	243.84	
Agreement	17.90	119.31	467.22	
Assertion	38.82	3.70	877.07	
Commissive	52.28	193.85	708.38	
Reassuring	38.82	286.39	1,129.99	
Total	994.31	3,276.92	13.747,08	
Deviation across groups				18,018.31

Table 7 The four most frequently employed expressive speech acts in each subcorpus

	Pr subcorpus	SL subcorpus	D&T subcorpus
Speech acts	Apology	Compliment	Thanking
	Compliment	Thanking	Wish
	Concern	Greeting	Greeting
	Greeting	Directive	Compliment

ranks among the top four in two of the subcorpora (SL and D&T), while Apology, the most frequent expressive in Pr, ranks fifth in both SL and D&T. These tendencies may well be accounted for by the students' focus on assuring a good rapport with the others, rather than on their own feelings. This other-focus is probably motivated to a certain extent by the blended nature of the learning context: students have other means of communication (not supervised by the lecturer). Consequently, it may be considered as a safe guess that, if they wished to vent their feelings about issues other than the online activities, they would probably resort to channels other than the e-forum, such as face-to-face conversations, email messages or Whatsapp.

The other-focus of the four expressive types mentioned above is enhanced in the data by the use of typographic signs like exclamation marks or emoticons, which seem to be preferred in other-directed acts rather than in self-directed acts such as Liking or Concern. Examples of use of these signs are the Thanking in (30), the Greeting in (31) or the Apology in (32):

(30) Thanks, Anat for offering to put the analysis in the final document! ☺ (SL)

(31) Hey guys! (D&T)

(32) Hi, sorry for being this late, I've been having problems with my internet connection at home ☹ (Pr)

It is to these other-oriented speech acts with a relatively high frequency of occurrence across the three subcorpora – Compliment, Greeting, Thanking and Apology – that the analysis will now turn.

4.2.1 Compliments

Compliments are relatively frequent in the three sets, in all cases ranking among the four most frequent expressive types (Table 7). In general terms, students employ them to positively evaluate their partners' previous work or to encourage group work, as illustrated by examples (33–35) taken from the three datasets:

(33) Perfect Cristina! (Pr)
(34) Kasia, good job with the colours, I really like it. ;-) (SL)
(35) Good work girls! (D&T)

The conventionalization of Compliments may be seen in the frequent use of a reduced number of adjectives such as *excellent, fine, good, great, interesting* and *perfect*.

(36) Great to hear that! (D&T)

The different ranking of Compliments across the three subcorpora (see Tables 3 and 7) suggests that their high frequency seems to be connected to linguistic proficiency: they rank slightly higher in the master students' subcorpus (21 %), and seem to gradually drop down along the proficiency scale, with 16.54 % in Pr – the undergraduate evening group with an older age and strong academic concerns– and 14.89 % in D&T, the morning large, mixed ability undergraduate group. In this group, precisely, Compliments are predominantly issued by the most proficient students, who seem to congratulate and encourage their mates' contributions more frequently than lower level students. These proficient students who tend to compliment other members in a lower level group could be taking on the role of facilitators and social mediators, very much as a teacher would.

In contrast, Compliments are pervasive in the contributions by the master's students, to the extent that they are the most repeated expressive act above other Expressives such as Thankings or Apologies. Given their nature, Compliments are highly suited to boost the good rapport between the interlocutors. Furthermore, the addressee's positive evaluation by the writer often serves also to emphasize or stand for other speech acts like Thankings (Maíz-Arévalo 2010), as shown by example (37):

(37) Thank you for your interesting comments, Anat! (SL)

This association between frequency of Compliments and proficiency might be related to the fact that Compliments, despite their apparent innocence, are rather complex speech acts (Maíz-Arévalo 2010), especially so when they may be serving

a hidden agenda. In other words, they can be used in combination with Directives as pre-requests as in example (38), where the writer flatters the addressee whilst implicating her future work:

(38) I'm sure u'll have many interesting points to add. (SL)

The pragmatic and linguistic relative complexity of these compliments can account for the master's students' comfort in using them, as opposed to the other two groups, whose compliments are more frequently connected to the genuine positive evaluation of previous work and, in general terms, might also be said to be linguistically easier to realize.

4.2.2 Greetings

Greetings are often typographically marked in the data, and rank among the four most frequent Expressives found in the three subcorpora (Table 7). However, their distribution is quite uneven, ranking highest in D&T (16.14 %) while they amount to 13.73 % in SL and 9.02 % in Pr. The predominance of Greetings in D&T may actually be explained by group size: with 64 students, many of the participants did not initially know one another nor did they even sit together in the classroom, so that the e-forum was often their only "meeting-place". In this context, emphasizing Greetings as Expressives may be interpreted as an attempt to open up and close the communication channel as "emotionally" as possible, building up a warm "hello" which seldom took place in face-to-face classroom communication, and trying to communicate willingness to collaborate and to build rapport. However, in the other two groups, Pr and SL, consisting of 9 and 10 students respectively, group members knew one another very well and were involved in daily face-to-face greetings; consequently, the need to use emphatic Expressives was probably seldom felt.

Group size, thus, emerges as one of the variables with a bearing on the differences in expressive speech act types found in the data, together with linguistic proficiency, in line with the tenant that, in Computer Mediated Communication, lack of non-verbal, kinetic information has to be made up for linguistically and typographically (Herring et al. 2013).

The linguistic realization of Greetings in the corpus is highly conventional. The most frequent expressions are *hello* and *hi* with exclamation marks, by themselves or followed by other words or expressions like terms of address, as in examples (39) and (40):

(39) Hi girls! (D&T)
(40) Hi Anat ☺ (SL)

These realizations undoubtedly make Greetings highly amenable expressive types in mixed-ability groups like D&T, as they are an easy and effective way to build rapport and keep communication fluent. In fact, in the master group (SL),

Greetings marked as Expressives are not only less frequent, but also more sophisticated, combining exclamations and emoticons, as in (41) and (42):

(41) Good morning! ☺ (SL)
(42) See you in class! :-) (SL)

Greetings are the least frequent in group Pr, in which the students were remarkably task-oriented. In many cases, these students' contributions focus only on the task, without any initial or final Greeting. This is illustrated by (43–45), which contain three contiguous contributions. This scarcity of Greetings is due to the fact that students do not feel the need to be especially polite to one another: the role of the e-forum is to get the task done in collaboration.

(43) PCIs: "he rushed out" – something was wrong
"Sampson began looking through them" – Sampson was in charge just one more
(44) Another could be? "we went up, I and one or two others" – they were curious"
(45) ¿¿¿"He made some odd sort of noise in his throat" – Something was wrong with him???"

4.2.3 Thankings

Thanking is another linguistically simple and socially rewarding Expressive. Therefore, it is no wonder that it also ranks quite high in the data, being the most frequent expressive type in D&T (19.91 %) and the second most frequent in SL (18.88 %), just below Compliments. The acts of thanking in the data invariably contain *thanks* or *thank you*, with the possible exception of (46), which lies in between Thanking and Compliment, and was eventually classified as Compliment:

(46) It has been a pleasure working with you:-D (SL)

Although quite homogeneously present in SL and D&T, Thanking expressions are remarkably lower in Pr (6.77 %), the evening undergraduate group. It should be reminded that this group also ranked very low in Greetings, so the reason for the low percentage of Thankings could be expected to be the same, namely, the fact that Pr is strongly task-oriented, and participants do not find it necessary to be expressing their thanks to one another all the time. However, if this was the case, the also task-oriented SL master group should also display a low presence of Thankings. But this is not the case, as shown above. Thus, if these two groups do not differ in task-orientation, high linguistic proficiency, or size – nine and ten students, respectively, as opposed to the sixty-four students in D&T – then the reason for the gap in the expression of Thankings has to be found elsewhere. In fact, one further contextual feature differentiating SL from the other two groups was multiculturality: while in Pr and D&T most students were Spaniards, SL was a multicultural master group with students from different nationalities using English as a lingua franca, and probably aware of the importance of pragmatic knowledge in

intercultural communication. Thus, despite being in a small group too, students in SL probably relied on the frequent expression of thanks in close resemblance to British English, in which Thanking is far more common than in Spanish.

Regarding Thanking, consequently, it could be argued that, while its simple linguistic realization – "Thanks" or "Thank you" in most cases, accompanied by emoticons and/or exclamation marks – would predict a noticeably higher use by the large mixed-ability group D&T, as is the case with Greeting, the students in the small SL master group issue thanks on a British English basis, due to awareness that English was being used in a lingua franca setting.

4.2.4 Apologies

Apologies are quite common in the three subcorpora, ranking first in the Pr group and fifth in the other two groups. This high frequency is not surprising: the collaborative nature of the task requires regularity and accuracy in individual contributions; consequently, students feel the need to give excuses when they do not consider that their collaboration is as satisfactory as it should be. The Apologies found in the data often concern past or future inability to contribute in a satisfactory way in the writer's own perception, often accompanied with a justification regarding the Internet connection or other personal reasons (47). In other cases, the apology was carried out in terms of limitations of knowledge or ability (48, 49 and 50) or a previous wrong answer (51), and may be accompanied by speech acts of Concern (52):

(47) Hi, sorry for being this late, I've been having problems with my internet connection at home ☹ (Pr)
(48) I'm sorry, but I thought there would be another meeting. (D&T)
(49) I am sorry for not replying to you Andrés, but I do not remember that much about entailments. (Pr)
(50) I'm sorry I wasn't able to answer before. (SL)
(51) I think "words" is a new local topic, I put it on another list sorry for confusion! (Pr)
(52) Sorry I am quite confused. (D&T)

The expression of Apologies displays a quite high degree of conventionalization, since they are most often expressed with *sorry*: the occurrences of *(I am) sorry* in the three subcorpora amount to 76. Therefore, *sorry* may be considered as a conventional way of expressing Apologies. Apologies without *sorry*, such as (53–54), are much less common.

(53) I realized that we don't have to refer to it, but just relate it to agreement :-) (SL)
(54) but I don't know what that implies... maybe I was wrong and it is marked ideational... (Pr)

The frequency of Apologies is particularly outstanding in the Pr group, probably due to the particularities of their situation. For this group, the task as a whole

counted for the final grade, as well as individual participation in it. The students were on average proficient, had planned to get high grades in the subject, and were conscious that their colleagues were in the same situation. Time was considered valuable, and inadequate contribution to the task was acutely perceived as face-threatening.

In certain cases, especially in the Pr group, Apologies prompt Reassuring speech acts, which aim at communicating acknowledgement or encouragement towards other group members. An example of an apology followed by reassurance is (55), in which Vicente feels uneasy about having to stop contributing to the e-forum and Cristina reassures him emphatically:

(55) Vicente: I am sorry Cristina because you have already done a lot of work and I fear that you will have to do the rest. I just hope T has it present when thinking about our marks, jeje. Thank you very much.
Cristina: Vicente!!! dont be silly!!! without your ideas I wouldnt have been able to work out many of the answers!!!!! Thank you for opening the forum and giving me an idea how to continue answering!!!! ☺ ☺☺

4.2.5 Further Comments on the Similarities Among the Subcorpora

As can be observed, one of the recurrent similarities is the high degree of conventionalization in the linguistic realization of the Expressives subtypes described above. This cannot be satisfactorily explained by limitations in the students’ knowledge of English, since their proficiency should allow them to be more creative in their use of language. A plausible reason, together with the formulaic nature of these expressions also among native speakers themselves, is that priority is being given to the performance of the task, and this performance needed quick and effective rapport building other-centredness. This factor, together with the informal relations among the students –even though teacher’s supervision was an inhibiting factor –may have contributed to the limited effort made about the choice of an expression for performing these highly frequent Expressives.

4.3 Differences Across the Subcorpora

Despite the similarities discussed above, the three subcorpora are far from homogeneous in many respects. For instance, expressions of Concern are overwhelmingly higher in Pr than in the other two subcorpora, as can be observed in Table 3. SL, on the other hand, displays a remarkably higher presence of Compliments, Directives, and Assertions, while ranking below average in Wish, Concern, and Reproach. Finally, D&T presents above average percentages in Greetings and Wishes, and is also high in Thankings and Reproaches, but has a relatively lower presence of Directives, Assertions, and Agreements typographically marked as

Expressives. The reasons for these differences may be connected to contextual features with a bearing on the linguistic choices made by participants regarding the expression of emotion and of a mental state of desire.

4.3.1 Directives

With regard to Directives, it is interesting to note that the master's students perform this non-prototypical expressive act more frequently than the two other groups (13.3 % as opposed to 7.52 and 4.4 %). Despite this difference, there is still a certain degree of conventionalization in the expression of Directives (with formulaic expressions such as "feel free to . . ." or "Let's. . ."), as well as a common tendency in the three groups to employ typographic means –e.g. repetition of exclamation or question marks, use of emoticons, etc. –to soften the face-threat their Directive may entail for the addressee. Examples (56–58) show this common pattern:

(56) Whatever, I'm finding quite difficult to find a good PCI, could you help me out with that fellas?? (Pr)

(57) Could you add them in this file with some pretty color. ;-) (SL)

(58) so, as soon as you can, please complete it and then I think you could upload the final document ☺ (D&T)

However, a major difference between the two undergraduate groups and the master's group is that the latter resorts more often to typographically enhanced Directives. In this group, Directives are one of the four most commonly used Expressives. The main function that these Directives perform is to ensure that the collective analysis moves forward smoothly, with the full group taking part in the discussion. Thus, it is rather common to find writers asking for their partners' assessment of previous contributions to the collective task at hand, like in (59):

(59) I'll be waiting for your comments! ;-) (SL)

Furthermore, as opposed to the other two groups, who tend to demand the addressee's opinion by resorting to direct questions (see example 56 above), the master's students opt for the set formula "let me know. . .", which appears 11 times and is always accompanied by emoticons to soften the imposition, as in (60)–(62) below:

(60) So, let me know if you have a different interpretation:D (SL)

(61) Let me know what you agree/disagree with:-D (SL)

(62) Please check it out and if there is any 'disagreement', please let me know. But I don't think there would be a big disagreement. ;-p (SL)

Interestingly enough, resorting to emoticons in the master's group (the group with the oldest age average in the data) runs contrary to our initial expectations that younger students might be using emoticons more often than older ones. However,

we must remember that this group was particularly prone to the pervasive use of emoticons across types of Expressives; an account of plausible reasons for this tendency is provided in the subsection on Assertions (4.3.2).

This high frequency of Directives in the master’s group seems to prove that, while also trying to keep a good rapport among their colleagues, the students seem to be rather task-oriented. This tendency towards task-orientation is also partially shared by their evening undergraduate counterparts, whose use of directives ranks fifth among all the subtypes and is significantly higher than that of the morning group. Such a tendency may be due to the fact that the Pr and the SL groups have a limited amount of free time in comparison with the morning students, and therefore need to maximize their efforts without wasting unnecessary time. Their main objective, then, is to ensure that the task at hand is successfully done without endangering the group’s good rapport.

4.3.2 Assertions

This strong task-orientation among the master’s students might also account for their use of expressive Assertions, their percentage being also slightly higher –if not strikingly so –when contrasted with the use of this kind of Assertions in the other two groups (6.86 % in the SL group as opposed to 3.01 % in Pr and 1.44 % in D&T). Their higher academic proficiency might also have a powerful say in the way they express their ideas with a strong conviction, as in examples (63)–(65):

(63) His role is minimized in 2005, and Cinderella has a more important role:D (SL)
(64) I have included this in the document I´ve attached (highlighted in pink):-D (SL)
(65) For question 2, I tried to summarize before the table. It seems logical to put words before the table. ;-) (SL)

As reflected by these random examples, a common strategy is to accompany Assertions with a final smiley emoticon, thus “enhancing” the message. In other words, the image serves to add an extra meaning that was not expressed by the purely textual part of the message. In these cases, the added nuance of meaning can be described as the tribute paid to interaction and rapport maintenance by avoiding what might be otherwise interpreted as a rough or arrogant expression of personal opinions on the writer’s part. Indeed, this might partially explain why the use of emoticons is markedly more common among the master’s students, who use emoticons up to 131 occasions –in contrast with the 16 and 32 times used by the evening and the morning undergraduates, respectively.

The question that arises is what might move the master’s students, who are also linguistically more proficient, to resort to these apparently simple typographic marks. Several factors might be at play, starting with the fact that these marks are

an extremely quick resource to go on with the task at hand while simultaneously performing the following crucial pragmatic functions:

(i) They help to boost and keep group rapport by adding a friendly tone to what might sound as rough and too transactional messages,
(ii) They help minimise imposition on the addressee when accompanying face-threatening acts, acting as a politeness strategy.

However, should this be the only reason, the evening group – also rather task-oriented – might employ the same kind of resource as the master's group, but they do not. In addition, it is difficult to determine why those students with less linguistic proficiency (i.e. the morning group of undergraduates) do not favour the use of emoticons either. In light of this, we firmly believe that two are the determining factors at play. On the one hand, the multicultural nature of the master's group might incline students to opt for a more "globally acknowledged" expression of good will such as emoticons rather than risking other linguistic expressions that might lead to intercultural pragmatic failure. We must remember that their concern with politeness was also shown to have an influence on their realizations of Thankings and Compliments, and on the use of emoticons with Directives. In fact, some of these students, asked informally a posteriori, acknowledged that they might have overused emoticons in an attempt to make their message sound "more friendly". On the other hand, the method of assessment might also be influencing students' behaviour in each of the groups. As already commented in the methodology section, the master's students knew that the e-forum was only relatively controlled by the teacher, whose presence was limited to providing help when so required. In contrast, both undergraduate groups had a strong presence of the respective teachers, which might have put students off employing more colloquial and informal ways of keeping rapport with their partners.

4.3.3 Wishes

Wishes are remarkably more frequent in D&T in comparison to the other two subcorpora: 17.20 % in D&T, 6.77 % in Pr and 3.43 % in SL. The linguistic expression of Wishes is quite highly conventionalized: the most frequent realization is the verb "hope", as in (66) and (67), and, to a lesser extent, "look forward to":

(66)	I hope everything is ok. :-D (SL)
(67)	I hope you can give me an idea and do it together (D&T)

The relative simplicity of this realization is probably at the root of the high preference that students in the largest, lowest level group show for these Expressives. But the mixed-ability nature of this group may also serve to explain the very need to use wishes with such frequency: despite the fact that D&T contained a reasonable number of academically skilled, linguistically proficient and motivationally engaged students, it also contained an equal number of students who might

well have found the task difficult. It should be reminded that this group combined 2nd, 3rd, and 4th year students in an optional subject, and that no doubt it could be expected that younger, less experienced members with fewer years of exposure to academic input would find it more difficult to cope with the task, and thus rely on “wishful thinking” to communicate willingness to collaborate.

Apart from the relatively simple realizations of Wish as “hope”, other instances involve higher elaboration, as can be observed in examples (68), (69), and (70):

(68) I would like to begin with a very brief comment regarding the first point we have to deal with. (Pr)

(69) Looking forward to your comments :D (SL)

(70) Have a wonderful bank holiday! (D&T)

4.3.4 Concerns

Concerns do not rank among the most frequent Expressives in the data, as shown in Table 3, but may be worth discussing due to the difference observed in Pr, where Concerns peak up to 10.53 %, with respect to the other two subcorpora, D&T and SL, with 2.93 and 1.71 %, respectively. As was stated above, Pr students belonged to a small group with a strong academic motivation, who usually take time off families and jobs to attend lessons and engage in a course. Under such circumstances, collaborative learning becomes a demanding social activity that all participants feel responsible for, so that the high frequency of this type of Expressives in the Pr group may well be due to students’ anxiety about getting the activity well done, and not interfering negatively with other participants’ work. These reasons are basically the same as those that favoured the high frequency of Apologies: we must not forget that these two speech acts point to a negative situation that creates anxiety, the difference lying in that Apologies convey a feeling of guilt towards others while Concerns simply express uneasiness. This anxiety is often typographically enhanced by means of repetition of question marks or exclamation marks (71); in other cases, students ease their minds by humoristically using a metaphor (72), onomatopoeias (73), or emoticons (74):

(71) What is the PCI of:look a the papers on his desk??? (Pr)

(72) The GCI’s and PCIs are lost in a remote island (Pr)

(73) This is driving me mad hahahahahahha” (Pr)

(74) This is by far the most difficult question ☺ (Pr)

Expressions of Concern are extremely rare in SL (1.71 %). One of them is presented in (75):

(75) I’m just a bit worried because the document is 10 pages (though it includes many tables) :-)... (SL)

In the lower level group, D&T, this speech act type usually includes other implicit speech acts, such as requests for help (76), or implicit Apology and justification, as in (77). In some cases, expressions of Concern are accompanied by directives, as in (78), but their most frequent realization in this subcorpus is an adjective of affect, usually “lost” or “confused”, as in (79):

(76) But then, I don’t know how to classify them!! (D&T)
(77) Luisa and I are a bit lost and insecure about our responses; we are in the second course and we have the same knowledge (D&T)
(78) Let’s do it together because I am a bit lost with it! (D&T)
(79) I’m a little confused about it (D&T)

In spite of their similarity to Apologies, Concerns do not display a high degree of conventionalization. We may even hint that writers seem to display some creativity when they wish to express that they are worried without feeling apologetic, as in (77) above and (80):

(80) Ok. perfect, I’ll try it again later, but maybe she wanted us to say it is anaphoric :S I don’t know. (Pr)

4.3.5 Reproaches

The expression of Reproach also deserves further attention, in spite of being relatively infrequent in the three subcorpora (5.55 % in D&T, 4.51 % in Pr and 0.8 % in SL). The SL master students seemed to try to avoid this speech act more keenly than their Pr and D&T counterparts, who occasionally resorted to Reproach as a means to re-conduct the uncollaborative behaviour of certain group members who took too long to upload their contributions, or who simply did not show up in the e-forum, in either way hindering and endangering the collective grade to be derived from the activity.

The master’s students’ avoidance of Reproach seems surprising at first sight, since it was the only group whose students were graded exclusively on the basis of their collectively produced document (and had been previously informed about that). The reasons may well lie in their concern for politeness, which was here reported to be also present in their realization of other speech acts; probably, multiculturality ruled out rudeness as a risk with unpredictable consequences. In fact, SL contains just two cases of Reproach, and only one of them is explicitly so (81):

(81) and I don’t know why you don’t answer (SL)

D&T displays some cases where impatience may turn an originally implicit directive into an explicit Directive, as in (82) and (83). Here Clara joined the activity a week later than the others, so that when she started, her group mates had already been collaborating for a few days. Moreover, she got in touch with a

Commissive, promising to join the group by uploading her answers promptly, but five days later there were no news from her. Example (82) is the first Reproach, containing an indirect Directive. The following day indirectness gave way to what might sound like utter rudeness (83):

(82) Now we have to wait until Clara uploads her answers and then she has to complete the shared document (D&T)

(83) Clara, what about your answers? Please upload them and complete the word document until 19th (D&T)

Reproaches in D&T are occasionally combined with subtle threats, usually involving the teacher, as in (84). On other occasions they contain an Apology, showing awareness of the face-threatening potential of the utterance (85). But it is also possible to find utterly direct Reproaches, as in (86):

(84) I hope that the other two participants of the group say something, if not... I think we must talk to T (D&T)

(85) Sorry but I have been waiting during 2 weeks so that someone could answer me and nobody did (D&T)

(86) The next time, be more responsible, please (D&T)

As for Pr, the smallest undergraduate Group, it must be noted that its Reproaches are usually less direct than in D&T and never involve nomination, as in (87).

(87) The other person did not contribute at all. (Pr)

Reproach thus emerges as a multifaceted expressive speech act whose choice may be influenced, as in previous cases, by several contextual variables, namely assessment criteria, group size and multiculturality. The former seems to increase pressure and favour impoliteness. The second, group size, could lower the threshold of FTA imposition, as online conflict may not necessarily have noticeable effects on non-existing face-to-face social interaction. On the contrary, multiculturality could make participants prime safety in an already unstable set of interactional rules drawing heavily on British politeness and indirectness as the model to be followed in ELF (English as a Lingua Franca) communication.

5 Conclusions and Suggestions for Further Research

The analysis carried out in this paper, which covered expressive speech acts in a corpus consisting of three subcorpora of e-forum history logs derived from online collaborative writing activities on different subjects in English linguistics, has shown that Expressives emerge as an influential factor on the participants' rapport building, despite the task-oriented nature of the interaction. The study uncovered common features to the three subcorpora: predominance of other-oriented over self-oriented

Expressives and, in particular, a high frequency of Compliments, Greetings, Thankings and Apologies; a high degree of conventionalization in the realization of these four subtypes as well as Wishes; and frequent cases of hybrid speech acts which provoked difficulties for classification.

The analysis also revealed remarkable differences across the subcorpora, in terms of frequency of use, concrete linguistic realizations of individual subtypes, and the use of emoticons and other typographic marks. The study suggests that these differences may be successfully accounted for by the influence of five variables, namely cultural homogeneity, age, linguistic proficiency, range of assessment method, and size of the group in which the students were having classroom instruction. An ANOVA test revealed the statistical significance of such variables. This suggests that the discourse produced during online written collaboration is deeply affected by the blended learning nature of an educational environment which combines online and face-to-face interaction.

Suggestions for further research on Expressives in task-oriented e-forums include qualitative studies, such as analyses of individual realizations of different Expressives or of frequent patterns into which Expressives may be combined among themselves and with other types of speech acts. Quantitative analyses of larger corpora may also focus on form, and, more concretely, on the most conventionalized realizations of Expressives signalled in this study.

References

Adolphs, S. (2008). *Corpus and context*. Amsterdam/Philadelphia: John Benjamins.

Austin, J. L. (1975 [1962]). *How to do things with words*. Oxford, New York: Oxford University Press.

Bach, K., & Harnish, R. M. (1979). *Linguistic communication and speech acts*. Cambridge, MA/London: MIT Press.

Blake, R. J. (2011). Current trends in online language learning. *Annual Review of Applied Linguistics, 31*, 19–35.

Cole, M. (2009). Using wiki technology to support student engagement: Lessons from the trenches. *Computers and Education, 52*, 141–146.

Dillenbourg, P. (1999). *Collaborative-learning: Cognitive and computational approaches*. Oxford: Elsevier.

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In P. Reiman & H. Spada (Eds.), *Learning in humans and machines: Towards an interdisciplinary learning science* (pp. 189–211). Oxford: Elsevier.

Givón, T. (1982). Evidentiality and epistemic space. *Studies in Language, 6*(1), 23–49.

Givón, T. (1984). *Syntax: A functional-typological introduction* (Vol. I). Amsterdam/Philadelphia: John Benjamins.

Givón, T. (1990). *Syntax: A functional-typological introduction* (Vol. II). Amsterdam/Philadelphia: John Benjamins.

Herring, S. C., Stein, D., & Virtanen, T. (2013). *Handbook of pragmatics of computer-mediated communication*. Berlin: Mouton.

Kessler, G., Bikowski, D., & Boggs, J. (2012). Collaborative writing among second language learners in academic web-based projects. *Language, Learning and Technology, 16*(1), 91–109.

Leeming, D. E., & Danino, N. (2012). Breaking barriers: A case study of culture and facebook usage. *Journal of Modern Languages and International Studies, 1*(1), 52–64.

Maíz Arévalo, C. (2010). Intercultural pragmatics: A contrastive analysis of compliments in English and Spanish. In M. Luisa Blanco Gómez & J. I. Marín Arrese (Eds.), *Discourse and communication: Cognitive and functional perspectives* (pp. 175–208). Madrid: Dykinson and Universidad Rey Juan Carlos.

Martin, J. R., & White, P. R. R. (2005). *The language of evaluation*. New York: Palgrave.

Neumann, D. L., & Hood, M. (2009). The effects of using a wiki on student engagement and learning of report writing skills in a university statistics course. *Australasian Journal of Educational Technology, 25*(5), 382–398.

Nöel, S., & Robert, J.-M. (2004). Empirical study on collaborative writing: What do co-authors do, use, and like? *Computer Supported Cooperative Work: The Journal of Collaborative Computing, 13*(1), 63–89.

Piaget, J. (1928). *The language and thought of the child*. New York: Harcourt.

Rühlemann, C. (2010). What can a corpus tell us about pragmatics? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 288–301). London: Routledge.

Searle, J. R. (1976). A classification of illocutionary acts. *Language in Society, 5*(1), 1–23.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Storch, N. (2011). Collaborative writing in L2 contexts: Processes, outcomes, and future directions. *Annual Review of Applied Linguistics, 31*(1), 275–288.

Thomas, J. (1995). *Meaning in interaction. An introduction to pragmatics*. London/New York: Longman.

Verschueren, J. (1999). *Understanding pragmatics*. London/New York: Arnold.

Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Weigand, E. (2010). *Dialogue – The mixed game*. Amsterdam/Philadelphia: John Benjamins.

White, P. R. R. (2003). *An introductory course in Appraisal analysis*. http://www.grammatics.com/appraisal. Accessed 16 Mar 2009.

Yus, F. (2011). *Cyberpragmatics: Internet-mediated communication in context*. Amsterdam: John Benjamins.

Interaction and Codability: A Multi-layered Analytical Approach to Discourse Markers in Teacher's Spoken Discourse

Shanru Yang

Abstract This chapter introduces a novel multi-layered analytical approach combining corpus linguistics (CL), conversation analysis (CA), and second language (L2) classroom modes analysis (Walsh, Investigating classroom discourse, Routledge, London/New York, 2006; Exploring classroom discourse: language in action, Routledge, London, 2011) for the investigation of discourse markers (henceforth DMs) in the spoken discourse of teachers. In response to the DMs' multifunctional nature (Jucker and Ziv, Discourse markers: introduction. In; Jucker AH, Ziv Y (eds) Discourse markers: descriptions and theory, John Benjamins B.V., Amsterdam, pp 1–12, 1998), it suggests an integrated approach to examine both the macro and micro contexts of DMs in teacher-led classroom interaction.

Keywords Discourse markers • Classroom discourse • EFL teacher talk • Multi-layered analytical approach • Conversation analysis • Corpus linguistics

1 Introduction

According to Carter and McCarthy (2006: 208), DMs can be defined as:

> words and phrases which function to link segments of the discourse to one another in ways which reflect choices of monitoring, organization and management exercised by the speaker.

Lexical items like *right*, *yeah*, *well*, *you know*, *okay* have received a great deal of attention from various research perspectives in the field of linguistics (Jucker and Ziv 1998; Fraser 1999; Müller 2005; Romero-Trillo 2012). Mostly referred to as DMs, such words do not only have grammatical roles in discourse but also work as

S. Yang (✉)
School of Education, Communication and Language Sciences,
Newcastle University, Newcastle, UK
e-mail: yshanru@gmail.com

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2,
DOI 10.1007/978-3-319-06007-1_13, © Springer International Publishing Switzerland 2014

effective interactional features, especially in spoken conversations (Schiffrin 1987; Maschler 1998; Fraser 1999).

The basic characteristics of DMs, as Jucker and Ziv (1998) notice, have two identifiable categories: *diagnostic* and *descriptive* features. Diagnostic features are essential for researchers to decide whether a linguistic item is a DM whilst descriptive features provide additional information. Schourup (1999) emphasises three main diagnostic features, namely connectivity, optionality and non-truth-conditionality, which has a wide acceptance among researchers (Lee-Goldman 2011). DMs' connectivity refers to the ability to relate to discourse or talk units. In addition, DMs are syntactically optional and semantically independent. They are loosely attached to the syntactic structure and have little or no propositional meaning. Other descriptive criteria include initiality, orality, and multi-categoriality. In other words, DMs often appear in sentence-initial position, bear the feature of being oral, and originate from different source of lexical forms like adverbs and conjunctions. The multiplicity of DMs has been recognised in cognitive, social and textual domains (Schiffrin 2001). Among these different perspectives, researchers often find it difficult to choose among various terminologies, characteristics, and classifications to explore the nature, role and functions of DMs in discourse (Fischer 2006; Romero-Trillo 2012).

There are a growing number of studies that focus on DMs' applications in institutional talk, including psychotherapeutic practices (Tay 2011), interviews (Trester 2009), hospitals (Haakana 2002), and university lectures (Schleef 2008). In classroom contexts, DMs are found as useful guideposts for pedagogical clarification and effective interaction (Dalle and Inglis 1990). As one essential characteristic of teacher talk, DMs serve as a lubricant in conversations to reduce understanding difficulties, incoherence, and social distance between teachers and students (Walsh 2006; Fung and Carter 2007; Grant 2010). Dalle and Inglis (1990) suggest the importance of including DMs as part of communicative techniques in teacher training programmes. Yet so far, compared to the extensive research on DMs in second language acquisition (SLA) (see for example Müller 2005; Polat 2011), the relationship between DMs and the efficacy of classroom interaction is still under-researched, particularly in teacher's spoken discourse.

Traditional approaches to investigating DMs include discourse coherence model (Schiffrin 1987), grammatical-pragmatics (Fraser 1999), relevance theory (Blakemore 1992), and systemic functional grammar (SFG) (Halliday and Hasan 1976). There are also other alternative methods for the investigation of DMs. For instance, a dynamic-interactional approach is proposed (see Frank-Job 2006) to view DMs as a developmental process of pragmaticalisation, which underlies the multi-functionality of DMs in meta-communication. In recent years, more studies have tried to broaden the spectrum of studies on DMs by including new domains and approaches, like for example cross-cultural pragmatics (Aijmer and Simon-Vandenbergen 2011).

In response to the dynamic nature of DMs, this chapter aims to portray their multiplicity in academic discourse in higher education, focusing on college-level (middle aged) EFL teacher talk in particular. A mixed method that combines CL, CA, and L2 modes analysis is introduced to account for the multifunctionality of

DMs that a purely quantitative or qualitative analysis would not sufficiently describe. There are three aspects that this chapter is interested in:

- The synergy and appropriateness of combining qualitative and quantitative methods as a powerful methodological tool to investigate classroom discourse;
- Investigating the range and variety of DMs used in college EFL teacher talk in China;
- Exploring the use and functions of DMs in talk-in-interaction.

2 DMs in Pedagogical Settings

Any classroom, as Walsh (2006: 4) states, is a "dynamic" context where a series of *events* take place involving teachers, learners, discourses, settings, and learning materials. Classroom interaction is central to both teaching and learning, which is largely determined by an asymmetric role of participants, goal-oriented activities, and institutional needs (De Fina 1997). Conversation in a classroom context is a process which functions to build social rapport between teachers and students as one discourse community.

In pedagogical settings, a large number of studies have focused on the acquisition of DMs by L2 learners. Most of the studies agree that DMs are important indicators for L2 learners' communicative competence and oral fluency. For instance, Müller (2005) compares different functions of four DMs *so, well, you know*, and *like* used by German EFL university learners and American native speakers (NS) based on a spoken corpus which contains 70 recorded conversations between students after they have finished a film and task in class. The results show that differences occur in the usage of the individual functions of DMs. Similarly, Romero-Trillo (2002) focuses on the pragmatic fossilization of DMs in both child and adult non-native speakers (NNS) in Spain during their learning English process. The study shows that the linguistic production of NNS has lower competence in pragmatic development of L2 language compared to NS. Dailey-O'Cain and Liebscher (2006) reveal a functional distribution and specialisation of DMs *so* and *also* in mixed code of German and English used by L2 students in bilingual classrooms. O'Keeffe et al (2011) further note that the use of DMs has high pragmatic value in interaction that can be designed in awareness building tasks in class for L2 learners.

Recent studies have attempted to unfold the patterns of DMs in teacher's spoken discourse (Schleef 2008; Othman 2010). Teacher talk, especially in the L2 classroom, shares great similarities with foreign talk (Henzl 1973) or caretaker talk. As Kumaravadivelu (2006: 67) states, it is:

> characterised by a slow rate of delivery, clear articulation, pauses, emphatic stress, exaggerated pronunciation, paraphrasing substitution of lexical items by synonyms, and omission, addition, and replacement of syntactic features.

The use and functions of DMs in teacher talk, however, have not been fully described in the literature. For instance, Othman (2010) investigates three specific DMs *okay*, *right* and *yeah* used by NS lecturers in Lancaster University, UK. It is

found that college lecturers use DMs as signposts on a structural level when taking turns in lectures. Though DMs' functional significance is recognised in the study, there is a lack of quantitative description of how the three DMs distribute differently. Other studies observe that DMs have a positive role in reducing the difficulties of L2 learners' listening comprehension in lectures (Flowerdew and Tarouza 1995; Jung 2003; Elder and Golombek 2003; Amador Moreno et al. 2006). Those beneficial effects of DMs on enhancing learning and interaction need to be addressed in language instructions, especially in teacher education.

Based on the theoretical models of Schiffrin (1987) and Maschler (1998), Fung and Carter (2007) propose a core functional paradigm which serves as a useful metalanguage to categorise DMs in pedagogical discourse. The study examines the use of DMs by NS and NNS on the basis of a comparative study of two pedagogical corpora, CANCODE (the Cambridge and Nottingham Corpus of Discourse in English) and natural transcripts of recordings taken from interactive classroom discourse of secondary pupils in Hong Kong. The characteristics which DMs contribute to discourse coherence include: *interpersonal, referential, structural* and *cognitive* categories. The *interpersonal* category correlates the relationship between the participants (*really, obviously, you know*). At *referential* level, DMs connect preceding and following segments in meaning (*because, thus,* but). In *structural* category, DMs function to signal connection and transition between topics (*now, ok, right, well, by the way, firstly, so, how about, to sum up*). DMs also work at *cognitive* level in mental construction including thinking process denote (*well, I think*) and reformulations (*I mean*, *in other words*). This functional paradigm of DMs is effective in that it provides a descriptive framework to analyse DMs at multi-dimensional levels. Though the framework focuses on the acquisition of DMs in NNS learners, it emphasises a context-based model to investigate DMs from a functional perspective in L2 classroom discourse, which can be further applied to investigate the use of DMs in teacher talk.

3 Synergy of Corpus Linguistics and Conversation Analysis

The synergy of CL and CA analyses has been applied in various contexts as a useful approach to look at the linguistic and conversational patterns that exist in various types of spoken discourse, such as survey designs (Campanelli et al. 1994), political interviews (Carter and McCarthy 2006), health care contexts (Adolphs et al. 2004), educational settings (Llinares-García and Romero-Trillo 2008).

As popular approaches to study spoken discourse, CL and CA have their own advantages and limitations (Walsh et al 2011). According to Aijmer (2002), the use of corpora provides the opportunity to study the distribution and function of DMs in extensive text extracts. By focusing on a relatively large scale of data, CL is often found to look at interaction from a higher level (Walsh et al 2011). On the other hand, CA analysis focuses on the microscopic details in the interactional organization of turn-taking, sequence and repair (Sacks et al. 1974). Though being criticised for its

lack of systemic analytical categories, fragmentary focus and mechanistic interpretation of conversation (Eggins and Slade 1997), CA unveils the small segments of interaction in a case-by-case manner which simply cannot be reached by CL (O'Keeffe and Walsh 2012). The following categories generalise the common and theoretical grounding shared by CL and CA:

- Common ground:
  - Data resources: both use empirical and naturalistic data from "real world" rather than intuitions;
  - Procedure: both data are from a principled collection of texts;
  - Analysis: both investigate actual patterns of "language in use" with its social contexts;
  - Focus: both probe into iterative development in language;
  - Reference: both allow baseline comparison with other registers (sequential order in CA and reference corpora in CL).
- Complementary ground:
  - Scope: CL allows sizable and traceable selected texts while CA provides limited yet detailed collections;
  - Recurrent significance: CL techniques of actual frequency, distribution and lexical choice supports conversation analysts in describing the regularity of recurrent orderliness;
  - Analysis: the systemic way that CL identifies and characterizes words with associated linguistic features complements CA analysis;
  - Perspective: CL provides macro linguistic patterns in bigger picture while CA emphases on micro-contexts.

From the above discussion, it can be argued that CLCA are compatible methodological approaches and, thus, can be incorporated in the iterative development of analytical framework and a comprehensive view of contexts.

4 Data

The data for the present study come from nine hours' naturalistic video recordings of Chinese college EFL classes, recorded originally as part of a 3-year research project "EFL Classroom Discourse Research and Teacher Development" (Ref. No. 07BYY036), supported by China's National Social Science Grant. The video recordings were collected from one Chinese university in 2009, Beijing, P.R. China. The total word count of the spoken corpus is about 60,000 (59,959) (see Table 1). In the database, six experienced EFL Chinese teachers (two males and four females) and 144 Chinese students have participated (24 students per class). All the participants including teachers and students are L2 learners. The students studied an English major or did a combined degree of international journalism and English. In terms of the class types that constitute the corpus, there are three intensive reading classes

Table 1 Research project "EFL classroom discourse research and teacher D\development"

Class type	Teacher	Duration (h)	Word count	Relative frequency (%)
Academic writing	A: Male, middle-aged	1.5	9536	15.9
	B: Female, middle-aged	1.5	10604	17.7
Oral debating	C: Female, 40s	1.5	3740	6.2
Intensive reading	D: Male, middle-aged	1.5	12,286	20.5
	E: Female, middle-aged	1.5	11,627	19.4
	F: Female, middle-aged	1.5	12,166	20.3
Total	6 teachers	9	59,959	100

Fig. 1 Snapshot of Chinese College EFL English classes (Photo Author's own)

(Teacher D, E, F) which takes up 60.2 % of the corpus, two academic writing classes (Teacher A and B, 33.6 %), and one oral debating class (Teacher C) which has the least percentage 6.2 %.

The following figure (Fig. 1) demonstrates a snapshot of an intensive reading class of Chinese college English classes.

5 The Approach

5.1 Mixed Method: A Multi-layered Analytical Approach

Mixed methods research, according to Dörnyei (2007: 24), "involves different combinations of qualitative and quantitative research either at the data collection or at the analysis levels". Mixed methods, not only provide complementary forms,

but also support each other by integrating embedded quantitative and qualitative data at different levels, which can then make the results "more meaningful" (Dörnyei 2007: 273).

This chapter presents a multi-perspective analytical framework to investigate DMs in L2 classroom teacher talk. The data is analysed using integrated mixed methods techniques, also called *mixed analyses* (Onwuegbuzie and Teddlie 2003: 352). A mixed method analysis allows simultaneous examination of data at both micro and macro levels (Bazeley 2009). It, thus, aims to provide a more comprehensive analysis by enhancing representation and legitimization of the data. The multi-layered analytical approach includes pedagogical, quantitative and micro-analytical layers by following a principled procedure to probe into different levels of discourse in terms of pedagogy (L2 classroom modes analysis), frequency and distribution (CL), and discursive patterns in talk-in-interaction (CA).

The pedagogical layer focuses on the use of DMs in micro-contexts of L2 classroom discourse. Developed by Walsh (2006, 2011), the Self-Evaluation of Teacher Talk (SETT) model is a useful metalanguage for portraying L2 classroom discourse on the basis of understanding pedagogic goals and interaction features. A mode, as Walsh (2006: 111) defines, is

> an L2 classroom micro-context which has a clearly defined pedagogical goal and distinctive interactional features determined largely by a teacher's use of language.

In the L2 classroom, there are four types of modes that can be identified: *managerial mode* where teachers' main task is to manage students' learning process, *materials mode* where classroom activities are constrained by the subjects or topics, *skills and system mode* where interaction between teachers and learners are mainly centred on language skills practices, and *classroom context mode* where students have more opportunity to participate in teacher-student interaction. The model of L2 classroom modes serves as an effective platform for researchers to investigate and reflect on classroom discourse (Walsh 2006, 2011).

The quantitative layer utilises corpus-based methods for the identification of the linguistic patterns of DMs in classroom interaction. It provides a general overview of DMs in terms of frequency, distribution, and lexical choice. The lexical and grammatical patterns of DMs are presented and compared across different L2 classroom micro-contexts. This stage of analysis provides a launching pad by "taking the pulse" of the preliminary findings and isolating a smaller set of patterns for the researchers to start with (Adolphs et al. 2004: 25).

The micro-analytical layer serves to examine DMs in "talk-in-interaction" using CA. For any analysis of a text and its context, the basic step to consider is to analyse the text "on its own terms" (Ten Have 2007: 58). According to Aijmer and Stenström (2005), approaches to discourse and interaction like discourse analysis (DA) and critical discourse analysis (CDA) have been criticised for bringing pre-assumptions of wider contextual features to the micro-analysis, whilst CA "does not want to take off from grand social and political conceptions, as exemplified in the idea of *unmotivated looking*" (Ten Have 2007: 58). Derived from ethnographical traditions, CA offers "fine-grained descriptions" of the recurrence of "order" in the organization

of turn-taking and sequence, with valuable implications for language teaching activities (McCarthy 1998: 20). Compared with other critical discourse approaches, CA is more concerned with local co-construction of interaction without presuppositions before analysis (Schegloff 1997; Wooffitt 2005).

The multi-layered analytical procedure assumes the perspective of full integration: integration throughout both analysis and interpretation (Greene et al. 1989). The step-by-step analytical procedure can be summarised as follows:

1. Establishing sub-corpora based on L2 classroom modes
2. Selecting a particular mode
3. Detecting DMs in L2 classroom modes
4. Detecting the functional categories of DMs
5. Annotating DMs with regards to their discursive mode and functions
6. Analysing DMs using CL techniques
7. Examining the co-text using CA techniques
8. Examining the interactional features and pedagogy using L2 modes analysis

From step 1 to step 4, DMs are defined and detected mode by mode. Step 5 provides a manual annotation process to tag DMs occurred in teacher's spoken discourse. The data then undergoes three levels of examinations in terms of quantity (step 6), talk-in-interaction (step 7) and pedagogy (step 8). In combination, the circulated three stages of analysis (step 6–8) offer a deeper insight into how DMs perform at multiple levels of discourse in classroom interaction. By bringing qualitative and qualitative results together, the multi-dimensional analysis provides different levels of analytical angles of DMs in detail. However, how to adopt a multi-layered analytical research design raises methodological challenges for theory and practice (Aijmer and Simon-Vandenbergen 2011).

5.2 A Multi-layered Annotation Process

In the data treatment phase, to correspond with the tripartite analysis, a multi-layered annotation process is introduced to help identify the existence of DMs, their position in different modes, and functional aspect in interaction. By tagging DMs manually at multiple levels in WordSmith Tools (Scott 2008), the researcher can calculate the frequency of DMs distributed in different micro-contexts.

Corpus annotation or mark-up is the practice of adding extra information to the raw data through a manual or automatic process (Leech 2004). Different types of annotation include part-of-speech (POS) annotation, phonetic annotation, and discourse annotation (McEnery and Wilson 1996). Problem-oriented annotation allows the researchers to invent and encode what they want to investigate. The addition of tags or labels enriches the original raw data by adding values for the research purposes and by making them easier to quantify (Sinclair 2004).

Multi-functionality, as Leech (2004) points out, is one of issues that most annotation processes encounter. As traditional corpus annotation on DMs is unable to adequately describe their complexity due to the unpredictability of language use,

this study chooses a manual annotation process. Though it is unlikely for either manual or computer annotation process to achieve 100 % accuracy, by following the appropriate guidelines and procedure, human analysts can make annotation decisions with high consistency. The multi-layered annotation process is hence proposed in the following procedure:

- Identifying DMs
- Identifying L2 classroom modes where DMs are situated (Label: Mn)
- Identifying DMs in functional categories (Label: Cn)

Each of these steps will be illustrated in detail in the following Sections.

5.2.1 Identifying DMs

Adapted from Brinton (1996), Fung (2003) and Müller (2005), a list of the diagnostic features that provide the crucial test for DMs can be summarised as follows:

- Lexis: multi-resources, fixed/short/small, micro-markers;
- Prosody: pause, intonation, stress, accompanied with non-word verbalization;
- Syntax: flexible in position, independent between clauses, detachable, turn-initial or stand-alone position in referential relation;
- Semantics: independent, optional, no effect on truth condition;
- Indexicality: anaphoric or cataphoric between discourse units.

DMs are independent linguistic entities both syntactically and semantically. In syntax, they are flexible, insofar as they can adopt any position within an utterance and they are detachable to the structure. DMs can be inserted in initial, internal, or final position (Brinton 1996; Müller 2005). DMs are detachable or optional in syntactic and semantic structures, in the sense that to remove a DM does not affect the grammaticality nor the content meaning of the utterance (truth condition) (Fraser 1988; Schourup 1999; Fung 2003; Müller 2005). DMs themselves lack semantic meaning compared to other content words (Schiffrin 1987). According to Schourup (1999: 232), DMs actually "display", "reinforce", or "clue" the intended interpretation rather than "creating" additional meaning. For instance, *all right* in (1a) is a DM and a non-DM in (1b).

(1) a. **All right**. That is the end.
b. Are you **all right**?

In (1a) *all right* is semantically and syntactically complete and, therefore, it is a DM. However, *all right* in (1b) is a complementiser that cannot be removed. The deictic properties of DMs bridge the speaker, hearer and context as one holistic unit, which distinguishes them in discourse context. They index the utterance to local discourse, as well as the participants to the global context (Schiffrin 1987). Take *anyway* for example (Carter and McCarthy 2006: 208–209).

(2) a. I didn't really need it but I bought it **anyway**.
b. B: Oh. Right.
A: But er yeah, **anyway**, we drove in the rain and the dark for eight hours.

In (2a), *anyway* is used as a concessive adjunct, therefore does not have discourse marking function. On the other hand, A's *anyway* in (2b) signals a return to the main narrative after a digression as a DM.

Though there are many marginal forms and controversial cases, a discussion of their diagnostic features may assist in the identification of DMs. As Schiffrin (1987) emphasises, additional features like discourse context is also part of the reasons that causes the complexity of DMs in the process of grammaticalization.

5.2.2 Identify L2 Classroom Modes

After defining DMs, the next step is to establish the micro-contexts where DMs are situated in sub-corpora. The corpus is divided into four sub-corpora in terms of L2 classroom modes where they occur, namely managerial mode (M1), materials mode (M2), skills and systems mode (M3), classroom context mode (M4). The principled rationale to detect four patterns of L2 classroom micro-contexts follows the CA mechanism (turn-taking system, sequential structure, topic management), in relation to interactional features and pedagogical goals (Walsh 2006). According to Walsh (2006: 65), each L2 classroom mode marks itself with unique "fingerprints" in terms of linguistic, interactional, and pedagogic features, which differentiate themselves from each other. The process of establishing sub-corpora by L2 classroom modes can be illustrated in excerpt 1 as follows:

Excerpt 1

```
    1  T:   There are people who are vegetarian by preference
    2       by preference and by religion we know that like
    3       Buddhists. They don't eat meat at all. (0.1) So,
->  4       that's paragraph three. What is the author doing?
    5       Here, right in this whole paragraph. Is he trying
    6       to tell us that you know there's something wrong
    7       with his education? In this paragraph?

    8  S17: It tells what education has taught him.

    9  T:   That is to get him in touch with those differences
   10       among different nations or cultures so here he's
   11       talking about his what his education had actually
   12       taught him right? Even, even, probably where his
   13       education was right was was helpful was useful
-> 14       right? Was useful. Okay. That's paragraph three.
   15       Now paragraph four, S18, can you please read this
-> 16       paragraph for us? Slowly, deliberately, and loudly.
```

In total, there are three classroom modes detected in the excerpt above, which can be summarised as classroom context mode (M4) (line 1–4), materials mode (M2) (line 4–14), and managerial mode (M1) (line 14–16). In extract 1, the teacher is discussing the concept of vegetarianism, a term that is featured in the class textbook. In line 4, by stating “so, that’s paragraph three”, the teacher finishes up the previous discussion (line 1–4) in classroom context mode. He then guides the students back to the material by continuing the turn to raise a new topic “what is the author doing” from the material (line 4–14). From line 14 to 16, the teacher is moving from materials mode to a new learning activity by asking S18 to read aloud (line 15–16), using transition markers like *okay* and *now* (line 14, 15). The managerial mode can be detected and marked from line 14 to 16. As Walsh (2006) states, though mostly found at the beginning of a lesson, managerial mode constantly occur in transition of different modes to link two adjacent learning stages.

5.2.3 Identifying the Functional Categories of DMs

More challenges are faced when detecting DMs in functional paradigms. Between different conversational action boundaries, DMs are found as part of the fingerprints to manifest those shifts of interactional moments (Maschler 2009). Goffman (1981) points out the shifts of frames of natural talk, or footing, constantly undergoes changes in the alignment of the speaker and the hearer. Maschler (2009), on the same wave length, argues that our frames for events often switch into other contextual realms during interaction.

As discussed previously in Sect. 2, four types of functional categories of DMs (referential, structural, interpersonal and cognitive category) are observed (Fung and Carter 2007). In order to represent different functions that DMs may perform, a label <Cn> is given to every DM with regards to the four different categories: referential category (C1), structural category (C2), interpersonal category (C3), cognitive category (C4). In addition, one multi-functional category (C5) is added when DMs perform multi-tasks or ambiguous functions. Excerpt 2 provides an example of annotating the functions of DMs.

Excerpt 2

```
1    T:   <C2>Okay. What are: the other reasons? Why people
2         are vegetarians?
3    S15: =I think they were uh they prefer to be a
4         veget'rian to live uh more healthy life.
5    T:   =<C5>So it's about health <C3>right?
```

In excerpt 2, the teacher draws on the opinions from the students to discuss the reasons for being a vegetarian. There are several functions that DMs perform here. The teacher uses *okay* (line 1) to mark a shift to a sub-topic of the main topic (therefore structural category C2) and later S15 produces *health* as one potential answer. *So* (line 5) then performs at multi-functional level (C5), in that it enables the teacher to gain the floor back (structural C2) and signals an alignment between

Table 2 Representation of the labels

Categories	Labels
Modes	Mn
Managerial mode	M1
Materials mode	M2
Skills and systems mode	M3
Classroom context mode	M4
Functional paradigm	Cn
Referential category	C1
Structural category	C2
Interpersonal category	C3
Cognitive category	C4
Multifunctional category	C5

Table 3 Combination of DMs in L2 classroom modes and functional paradigm

MnCn		L2 classroom modes: Mn			
		Managerial	Materials	Skills and systems	Classroom context
Functional paradigm:	Cn	M1	M2	M3	M4
Referential	C1	M1C1	M2C1	M3C1	M4C1
Structural	C2	M1C2	M2C2	M3C2	M4C2
Interpersonal	C3	M1C3	M2C3	M3C3	M4C3
Cognitive	C4	M1C4	M2C4	M3C4	M4C4
Multifunctional	C5	M1C5	M2C5	M3C5	M4C5

the participants (interpersonal C3). In line 5, the DM *right?* with a question mark clearly marks an interpersonal category (C3) by seeking for a confirmation from the student (Beach 1993).

5.2.4 Presenting Corpora with Multi-layered Annotation

The final stage of compiling corpora is the application and presentation after data collecting, transcribing and coding process. This is a stage that heavily relies on the software utilised (Knight 2009). There are two types of software used in my analysis to represent the data in this study, namely WordSmith Tools (Scott 2008) and Transana (Fassnacht 2012). In WordSmith Tools, by inserting <symbol> tag in concordance analysis can detect all the DMs occurred in the spoken discourse. The following representation (Table 2) illustrates the categories, and their corresponded labels.

The column on the right represents the multi-layered coding symbols: Mn = L2 classroom modes, Cn = functional paradigm that DMs operate in. The column on the left is the meaning of each code. In the data, DM is represented and annotated as <MnCn> DM, with two types of tags. The number of the actual DMs can be calculated by concordancing on <M*C*> in WordSmith Tools, in order to exclude all the non-DMs. Table 3 displays an exhaustive list of the combination of DMs in modes and functional categories.

T: <M2C2>Okay, the whole picture of the world. That's what my education did for me. I had some <M2C3>kind of general understanding of how (.) the world (.) worked or how the world looked, uh (0.1). <M2C2>On the whole, "it taught me how to recognize easily and instantly the things that make one place or one people different from another." <M2C2>So we have two words here: two adverbs, "easily" and "instantly". Remember quick fix society?

Ss: Yeah.

T: What is quick fix society about? Makes things easy and (.) <M2C3>probably instant and fast <M2C3>right? You can get that uh fast. Uh (0.1) even though it's superficial. <M2C3>probably (0.1) uh, he's also talking about the superficiality. SUPERFICIALITY? The noun of, noun of superficial <M2C3>right? "easily and instantly the things that make one place or one people from (.) another." He already have the expression "one (.) people" <M3C1>and (.) this has something to do with the usage of the word "people" [S: people]. <M3C1>Because we know, "people" is usually used as a collective noun. Remember collective noun? Collective noun uh is a noun which refer to a group of things, a number of things, not <M3C3>just a one (.) single (.) thing, <M3C3>right? Why here "one people"? Why here "one people"? (0.1) S5.

Fig. 2 Sample of the multi-layered annotated transcription (Photo Author's own)

Concord

File Edit View Compute Settings Windows Help

N	Concordance	Set	Tag	Word #	Sen	Sen	Para	Para	Head
1	and Africa? We've talked about that <M1> right? And then what makes the			197	11	33%	0	1%	
2	sure you still remember this word "test" <M1> right? Remember Asia and			188	9	91%	0	0%	
3	Africa? We've talked about that right? <M1> And then what makes the			198	12	6%	0	1%	
4	, our TODAY, not his TODAY All right? <M1> And then what time does it			377	32	2%	0	1%	
5	, our TODAY, not his TODAY All <M1> right? And then what time does it			376	31	39%	0	1%	
6	T: okay, let's get started (0.1) <M1> uh (.) well, we got (.) got to the			5	0	19%	0	0%	
7	T: <M1> okay, let's get started (0.1) uh (.)			0	0	0%	0	0%	
8	T: okay, let's get started (0.1) uh (.) <M1> well, we got (.) got to the			7	0	26%	0	0%	
9	our discussion of this uh lesson. <M1> uh, (.) we got to the first, second			38	1	33%	0	0%	

Fig. 3 Output of concordancing on tagging in WordSmith tools (Photo Author's own)

Figure 2 shows a sample of the multi-layered annotated transcription. Any distribution of DMs in modes or functional categories is obtainable through searching relevant tags. For instance, the total amount of DM operating at referential level (C1) in managerial mode (M1) is accessible by searching tag <M1C1> in concordance analysis.

An example of the output using concordancing on multi-layered tagging in WordSmith Tools can be seen in Fig. 3. The concordance search on tags allows the users to track and count the statistical patterns of DMs in teacher's spoken discourse. For example, in Fig. 3, by inputting tag < M1 > search, the overall number of DMs existed in managerial mode (M1), together with the concordance lines can be accessed immediately.

After retrieving the frequency and distribution of DMs through CL, the transcription with multi-layered annotation then is inputted into Transana, computer-assisted software for qualitative analysis. Figure 4 is a snapshot of time-stamped

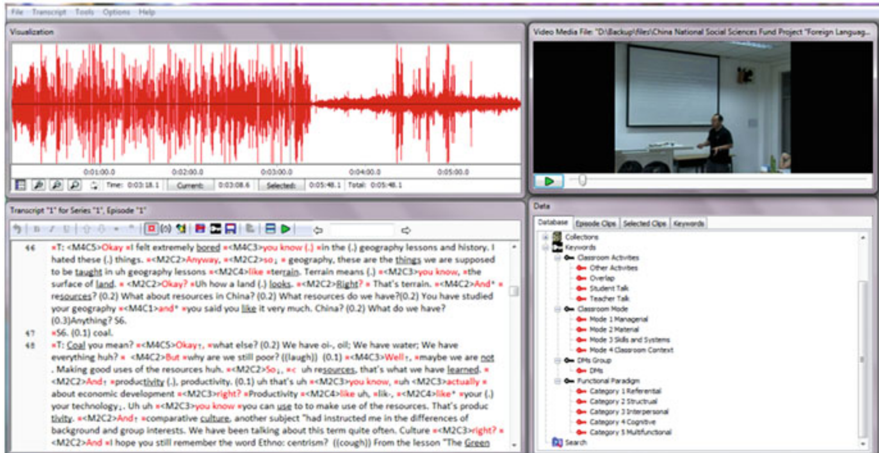


Fig. 4 Presentation of transcription in Transana (Photo Author's own)

transcription with multi-layered annotation. The interactive window of Transana not only allows the transcriber to edit their datasets with transcription conversations and notes but also synchronises the multi-transcripts and the video by adding time stamp and key word themes (Knight 2009).

6 Data Analysis

6.1 Analysis of the Data from the Corpus Linguistics Perspective

In terms of the frequency and distribution of DMs, there are 5,187 DMs discovered in Chinese college EFL teacher talk. Among the nine-hour video recordings (60,000 words), the overall teacher talking time (TTT) is about 417 min (77 %), of which intensive reading class takes up 89 %, academic writing class 80 %, and oral debating class 37 %). Frequency analysis shows that DMs have a significant contribution to contribute TTT (10 %) as well as to the whole spoken corpus (9 %). On average, a Chinese college EFL teacher produces 12.4 DMs per minute in class in the spoken corpus. This finding accords with Maschler's (1998) early observation, which shows a similar rate of occurrences in Israeli Hebrew casual conversations (roughly one DM every 4–5 s).

Figure 5 shows that the distributive trend of DMs in effect corresponds with L2 classroom modes. In other words, there is a positive correlation between the utterances Chinese college EFL teachers use in class and the production of DMs in their talk. In Fig. 5, discourse is used in the materials mode most frequently in terms of word count (24,195, 40.4 %), followed by skills and systems mode (16,991, 28.3 %) and classroom context mode (10,776, 18 %), with the managerial

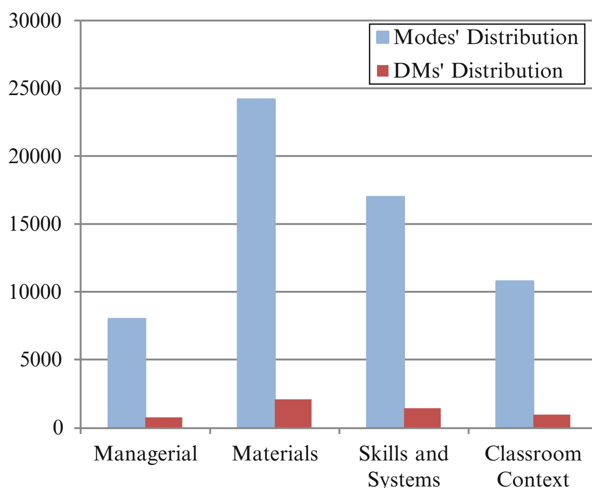


Fig. 5 Distribution of L2 classroom modes and DMs in Chinese College English classes

mode being the least frequently used (7,997, 13.3 %). Accordingly, DMs distributed in discourse in the materials mode occurs most often in this data (2,073) accounting for approximately 40 % of overall DMs used in this data. DMs in skills and systems mode have the second highest percentage with 1,416 occurrences (27 %). DMs in the classroom context mode accounts for 943 (18 %) and managerial mode has the lowest occurrences of DMs (755, 15 %).

Though frequency counts of the modes and DMs both demonstrate a high distribution in materials mode compared to other modes, it is necessary to probe into the proportion that DMs contribute to each mode (Fig. 6). Interestingly, despite the fact that managerial mode has the lowest number of occurrences (13 %), it has the highest constitution of DMs which takes about 9.4 % in total. Classroom context mode also has a high percentage of DMs at about 8.8 %, which ranks the second. Materials mode and skills and systems mode share similar percentages in that the former contains about 8.5 % and the latter 8.3 %.

In the general CL analysis of the spoken corpus, the examination of DMs across different L2 classroom modes suggests that there is a possible relationship between the use of DMs and different L2 classroom modes. Of the two modes with different speech-exchange systems in terms of interactional features and organization, the managerial mode is patterned with extended teacher procedure talk and instructions while classroom context mode has the least teacher turns to encourage content-centred interaction (Kasper 1985). The fact that DMs exist mostly in the managerial and classroom context mode reveals that DMs are largely used by Chinese college EFL teacher in activities that centre on classroom management and the promotion of learners' L2 production. This observation confirms the findings by Walsh (2006), who argues that DMs are one of the canonical interactional features that particularly constitute teacher's procedure talk in managerial mode (See Sect. 6.2 for CA discussion).

The distribution of DMs in functional paradigm highlights the interpersonal function that DMs perform across the four L2 classroom modes (Fig. 7). Among

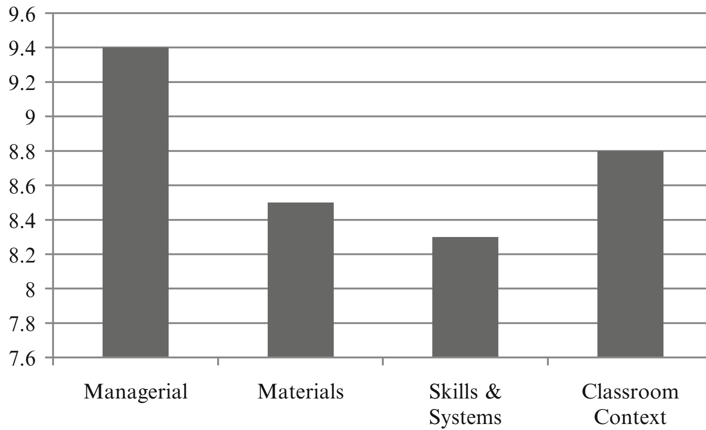


Fig. 6 Percentage of DMs that constitute each L2 classroom mode

Structural 539
Referential 403
Interpersonal 873
Multi-functional 138
Materials Mode 2073
Cognitive 120
Interpersonal 264
Referential 203
Managerial Mode 755
Structural 191
Multi-functional 74
Cognitive 23
Total DMs in 60,000-word corpus 5187 Chinese College EFL Teacher Talk
Interpersonal 595
Structural 343
Skills & Systems Mode 1416
Cognitive 88
Referential 245
Multi-functional 145
Cognitive 78
Classroom Context Mode 943
Interpersonal 448
Multi-functional 89
Structural 195
Referential 133

Fig. 7 Distribution of DMs in L2 classroom modes and functional categories

the 5,187 occurrences of DMs, 42 % (2180) are located in interpersonal category, which is the most frequently used function. Interestingly, DMs' high distribution in interpersonal category can also be found in all the four L2 classroom modes. Structural function is the second highest with 1,268 occurrences (24.5 %), followed by referential (19 %) and multi-functional category (14.5 %). The least frequent DMs are found in cognitive (C4) process (6 %).

All in all, the results from CL analysis provide a general overview of the statistical patterns of DMs across different L2 classroom modes. These results clearly demonstrate how the use of DMs in teacher talk can be affected by different interactional organizations of L2 classrooms, which intuition cannot sufficiently describe. CL analysis reveals that there is a reflexive relationship between teachers' use of DMs, classroom interaction and pedagogical purposes. As the classroom mode and pedagogical purpose vary, so does the distribution of DMs. The distributive trend of L2 classroom modes corresponds with that of DMs. A closer examination of their distribution in different modes suggests that though materials mode and skills and systems mode have the two highest occurrences of DMs, there appears to be higher percentages of DMs that constitute managerial mode and classroom context mode. This is summarised in Fig. 7.

6.2 Analysis of the Data from the Conversation Analysis Perspective

As CL has provided an overview of DMs in Chinese college EFL teacher talk, this section explores a micro-analytical perspective towards DMs in classroom interaction. It takes managerial and classroom context mode as two examples to demonstrate how teachers' use of DMs vary depending on different interactional organizations and pedagogical purposes.

6.2.1 Managerial Mode

As previous Fig. 6 shows, DMs constitute about 9.4 % of the managerial mode, which ranks the highest among the four modes. Featured with extended teacher turn and use of transition markers, the managerial mode often occurs at the beginning, transition or end of a university lecture (Walsh 2006, 2011). The teacher's main activity in this mode is to manage learning through formulaic language to transmit procedural information. Rather than monologic talk, procedural talk in effect involves teacher's awareness of the audiences, and therefore needs to be viewed as a jointly constructed process (Seedhouse 2004).

In the managerial mode, DMs frequently appear at the beginning, closing of extended teacher turns, and particularly at transitional moments between different classroom activities. Table 4 characterises the pattern of DMs in managerial mode.

Table 4 Pattern of DMs in managerial mode

Position	Pattern	DMs functions
Opening	Turn-prefaced DMs + instruction + pre-closing DMs	Turn-prefaced DMs: instruction initiator and attention getter
Transition		Pre-closing DMs: instruction finaliser and assurance seeker
Closing		

As Table 4 suggests, two types of DMs namely turn-prefaced and pre-closing DMs are identified in managerial mode. In the opening of a lesson, multiple DMs often occur at turn initial position to signal the coming of a new stage and to draw on the students' attention. Pre-closing DMs are particularly favoured by teachers to close up the lesson. Examples include a multiple use of tag-positioned DMs (e.g. *okay?*) to check students' progress and signal a completion of the lesson at the same time. When the managerial mode situates in closing position, there are no new actions/topics following. Excerpt 3 is an example taken from a transitional moment of an oral debate in class. The teacher in the excerpt is trying to organize the class to move on to the second round of debating practice.

Excerpt 3

```
((Ss discuss))
1   T:   ((looks around the room))↑all right are you ready for
2        the second uh round of practice? (1)((looks around
3        the room ))↑okay now (.) change: uh change roles. (.)
4        to the second students reading (12.) of the argument
5        (.) okay? (.)((looks around the room)) now(.)↑start
6        reading the argument ((looks around the classroom))
```

In this excerpt, DM *all right* (line 1) with a rise in tone interrupts the current actifity (group discussion) and prefaces the query of checking the students' learning progress. *Okay* and *now* (line 3) project the first instruction of changing roles, followed by a confirmation check *okay?* (line 5) in the end. The teacher looks around the room before initiating the two progressive signals. DMs *okay?* and *now* (line 5) as the pre-closing device terminate the instructional informing. The associated act of looking around together with marked prosodic features of rising intonation, pause and stress is constantly deployed by the teacher to get the students' attention and check whether the whole class is altogether. Finally her emphasis of *now* signals that this activity is happening at this moment.

To sum up, in managerial mode which consists of metastatement or in other words "saying-in-so-many-words-what-we-are-doing" (Garfinkel and Sacks 1970: 351), DMs play an important part as punctuation marks to help the learners to navigate their way, particularly in lecture comprehension (Breen 1998). In this mode, DMs often accompany with teacher instructions following the format of *turn-prefaced DMs + instruction + pre-closing DMs*. The next section chooses classroom context mode that centres on meaning-and-fluency.

6.2.2 Classroom Context Mode

When the teacher's aim is to maximise the opportunities for interaction, the pedagogical focus then shifts to the expression of personal meaning and the promotion of fluency (Seedhouse 2004). In contrast to the managerial mode, the management of turns and topics in classroom context mode is determined by the local context of speech (Walsh 2006). There is a major change towards a less narrow and rigid interactional organization affected by the pedagogical aims in this mode. In this mode, the role that the teacher plays is less prominent to ensure that sufficient space is allocated to the learners to manage and develop the interaction. The use of DMs is therefore likened to more naturally occurring conversation in terms of sequence organization.

In the classroom context mode where meaning-and-fluency is the focus, free-standing DMs with upward tones (e.g. ↑*okay* and ↑*yeah*) occur frequently as turn component unit (TCU) to signal the recipients' recognition and management of interpersonal relations. In this mode, stand-alone TCU DMs are used as minimal responses to mark active listenership desiring the learner's floor-holding to continue (McCarthy 2003; Knight 2009). Excerpt 4 is taken from an academic writing class. In this excerpt, the teacher is asking students to comments on a sample of writing in class.

Excerpt 4

```
1  T:   okay (.) very interesting uh any ↑question (1) any
2       question (3) any ↑question or any comment (1)
3       ↑comment (1) any ↑comment (.) yes

4  S3:  there is some- some problem

5  T:   ↑okay

6  S3:  uh with uh their classification

7  T:   ↑okay

8  S3:  of animals uh the insects is not the uh the- the- I
9       mean the (.) the the standard is not consistent

10      (.)

11 T:   ↑mhm how
```

As excerpt 4 shows, DMs are constantly used in classroom context mode as bridges between units. Upward-toned o*kay* (line 5, 7) appears as free standing to signal passive recipiency by working to retain the floor. DMs do not occur just anywhere but at the boundaries of TCUs to demonstrate that one unit has been received and that another is now awaited. Learners therefore treat it as a signal to continue (Beach 1995).

To sum up, in classroom context mode, which has the highest percentage of DMs accumulated in interpersonal category (47.5 %, Fig. 7), one reoccurring pattern discovered is upward-toned DMs as free-standing TCU in minimal responses. They serve as acknowledgement and floor-yielding tokens linking to active listenership.

7 Discussion

In response to the multifunctional nature of DMs, this chapter proposes a multi-layered analytical approach by combining L2 classroom modes analysis (Walsh 2006), with techniques used in CL and CA. The establishment of L2 classroom modes (Walsh 2006, 2011) and functional paradigm (Fung and Carter 2007) does not tend to generalise and encode interaction patterns, rather serving as a metalanguage to understand DMs in L2 classroom discourse.

Mixed methods data analyses can enhance the research findings by gaining deeper insights about the phenomenon in terms of data representation and legitimation. In this chapter, the synergy of CLCA has provided a comprehensive description of spoken interaction (O'Keeffe and Walsh 2012). According to Arminen (2005: 26), "CA and quantitative analysis do not mainly contradict each other; they simply address different orders of things". Though some theorists may argue that CA does not develop arguments on the basis of frequency data, it is in effect originally a reaction to the quantitative techniques in sociology which results in a strict empirical approach (Aijmer and Stenström 2005).

The multi-layered analytical approach looks at the ways in which DMs are used and the functions they perform in EFL teacher spoken language in higher education classroom discourse. CL analysis in Sect. 6.1 reveals that on average, a Chinese college EFL teacher produces 12.4 DMs per minute in class. Compared to other forms of discourse, the appearance of DMs in teacher's spoken discourse is remarkable. In addition, CL analysis highlights the high occurrence of DMs at interpersonal level in classroom interaction. CA analysis in Sect. 6.2 further explores the micro-contexts of DMs by analysing the examples from managerial mode and classroom context mode. It shows that teachers' use of DMs vary according to different interactional organizations and pedagogical goals. These results are only a representation of how DMs are used by Chinese college EFL teachers in the L2 classroom. Various social constrains like individual repertoires, social distances, speech types, and contexts may also affect the use of DMs (Schleef 2008).

8 Conclusion

Using nine-hour video-recorded Chinese college English classes, this chapter attempts to reveal that there is a reflexive relationship between teachers' use of DMs, classroom interaction, and pedagogical purposes. It proposes a multi-layered

analytical approach to unfold the complexity of DMs in the L2 classroom interaction. Methodically, it presents both the advantages and challenges of using a multiple analysis to investigate DMs in practice. This chapter hopes to shed some light on L2 teacher education, specifically in its potential to help teachers achieve their pedagogical goals.

Acknowledgements I am grateful to Professor Lian Zhang, who has authorised me to use the data from China's National Social Science Fund Project "EFL Classroom Discourse Research and Teacher Development" (Ref. No. 07BYY036). My special thanks to Professor Steve Walsh, Dr. Dawn Knight, Dr. Jesús Romero-Trillo, and Alastair Krzyzosiak for their advice and support.

References

Adolphs, S., Crawford, P., Brown, B., Sahota, O., & Carter, R. A. (2004). Applying corpus linguistics in a health care context. *International Journal of Applied Linguistics, 1*(1), 44–49.

Aijmer, K. (2002). *English discourse particles: Evidence from a corpus*. Amsterdam: John Benjamins.

Aijmer, K., & Simon-Vandenbergen, A.-M. (2011). Pragmatic markers. In J. Zienkowski, J.-O. Östman, & J. Verschueren (Eds.), *Discursive pragmatics* (pp. 223–247). Amsterdam: John Benjamins.

Aijmer, K., & Stenström, A. B. (2005). Approaches to spoken interaction. *Journal of Pragmatics, 37*, 1743–1751.

Amador Moreno, C. P., O'Riordan, S., & Chambers, A. (2006). Integrating a corpus of classroom discourse in language teacher education: The case of discourse markers. *Recall, 18*(1), 83–104.

Arminen, I. (2005). *Institutional interaction: Studies of talk at work*. Surrey: Ashgate.

Bazeley, P. (2009). Integrating analyses in mixed methods research. *Journal of Mixed Methods Research, 3*(3), 203–207.

Beach, W. (1993). Transitional regularities for 'casual' "okay" usages. *Journal of Pragmatics, 19*, 325–352.

Beach, W. (1995). Preserving and constraining options: "Okays" and 'official' priorities in medical interviews. In G. H. Morris & R. J. Chenail (Eds.), *The talk of the clinic: Explorations in the analysis of medial and therapeutic discourse* (pp. 259–289). Hillsdale: Lawrence Erlbaum.

Blakemore, D. (1992). *Understanding utterances*. Oxford: Blackwell.

Breen, M. P. (1998). Navigating the discourse: On what is learned in the language classroom. In W. A. Renandya & G. M. Jacobs (Eds.), *Learners and language learning* (pp. 115–143). Singapore: SEAMO Regional Language Centre.

Brinton, L. (1996). *Pragmatic markers in English: Grammaticalization and discourse functions*. Berlin: Mouton.

Campanelli, P., Channell, J., McAulay, L., Renouf, A., & Thomas, R. (1994). *Training: An exploration of the word and the concept with an analysis of the implications for survey design* (Research Series No. 30). Sheffield: Employment Department.

Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English*. Cambridge: Cambridge University Press.

Dailey-O'Cain, J., & Liebscher, G. (2006). Language learners' use of discourse markers as evidence for a mixed code. *International Journal of Bilingualism, 10*(1), 89–109.

Dalle, T., & Inglis, M. (1990). ITA "*teacher talk*" – *Discourse markers as guideposts* to learning (ERIC Document Reproduction Service No. ED 353 827). Washington, D.C.: ERIC.

De Fina, A. (1997). An analysis of Spanish *bien* as a marker of classroom management in teacher student interaction. *Journal of Pragmatics, 28*, 337–354.
Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.
Eggins, S., & Slade, D. (1997). *Analysing casual conversation*. London: Cassell.
Elder, C., & Golombek, P. (2003). The effects of organization markers on ESL learners' text understanding. *TESOL Quarterly, 37*(4), 749–759.
Fassnacht, C. (2012). *Transana* (Version 2.52). Wisconsin: Wisconsin Centre for Education Research.
Fischer, K. (2006). *Approaches to discourse particles*. Amsterdam: Elsevier.
Flowerdew, J., & Tarouza, S. (1995). The effect of discourse markers on second language lecture comprehension. *Journal of Studies in Second Language Acquisition, 17*(4), 435–458.
Frank-Job, B. (2006). A dynamic-interactional approach to discourse markers. In K. Fischer (Ed.), *Approaches to discourse particles* (pp. 395–413). Amsterdam: Elsevier.
Fraser, B. (1988). Types of English discourse markers. *Acta Linguistica Hungaria, 38*, 19–33.
Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics, 31*, 931–952.
Fung, L. (2003). *The use and teaching of discourse markers in Hong Kong: Students' production and teachers' perspectives*. Unpublished doctoral dissertation, University of Nottingham, Nottingham.
Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogical settings. *Applied Linguistics, 28*(3), 410–439.
Garfinkel, H., & Sacks, H. (1970). On formal structures of practical action. In J. C. McKinney & E. A. Tiryakian (Eds.), *Theoretical sociology: Perspectives and developments* (pp. 338–366). New York: Appleton.
Goffman, E. (1981). *Forms of talk*. Oxford: Blackwell.
Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed – Method evaluation designs. *Educational Evaluation Policy Analysis, 11*, 255–274.
Grant, L. E. (2010). A corpus comparison of the use of *I don't know* by British and New Zealand speakers. *Journal of Pragmatics, 42*, 2282–2296.
Halliday, M. A. K., & Hasan, R. H. (1976). *Cohesion in English*. London: Longman.
Haakana, M. (2002). Laughter in medical interaction: From quantification to analysis, and back. *Journal of Sociolinguistics, 6*(2), 207–235.
Henzl, V. (1973). Linguistic register of foreign language instruction. *Language Learning, 23*, 207–222.
Jucker, A. H., & Ziv, Y. (1998). Discourse markers: Introduction. In A. H. Jucker & Y. Ziv (Eds.), *Discourse markers: Descriptions and theory* (pp. 1–12). Amsterdam: John Benjamins B.V.
Jung, H. E. (2003). The role of discourse signalling cues in second language listening comprehension. *Modern Language Journal, 87*(4), 562–576.
Kasper, G. (1985). Repair in foreign language teaching. *Studies in Second Language Acquisition, 7*(2), 200–215.
Knight, D. (2009). *A multi-modal corpus approach to the analysis of backchanneling behaviour*. Unpublished PhD thesis, University of Nottingham, Nottingham, UK.
Kumaravadivelu, B. (2006). *Understanding language teaching: From method to postmethod*. Mahwah: Lawrence Erlbaum.
Lee-Goldman, R. (2011). *No* as a discourse marker. *Journal of Pragmatics, 43*(10), 2627–2649.
Leech, G. (2004). Recent grammatical change in English: Data, description, theory. In K. Aijmer & B. Altenberg (Eds.), *Advances in corpus linguistics: Papers from the 23rd international conference on English language research on computerized corpora (ICAME 23)* (pp. 61–81). Amsterdam: Rodopi.
Llinares-García, A., & Romero-Trillo, J. (2008). The pragmatic role of discourse markers of native and non-native teachers in CLIL contexts. In J. Romero-Trillo (Ed.), *Pragmatics and corpus linguistics: A mutualistic entente* (pp. 191–204). Berlin/New York: Mouton de Gruyter.
Maschler, Y. (1998). Rotsèlishmoakéta? Wanna hear something weird/funny? [lit. a segment]: Segmenting Israeli Hebrew talk-in-action. In A. H. Jucker & Y. Ziv (Eds.), *Discourse markers: Descriptions and theory* (pp. 13–59). Amsterdam: John Benjamins B.V.

Maschler, Y. (2009). *Metalanguage in interaction: Hebrew discourse markers*. Amsterdam/ Philadelphia: John Benjamins.
McCarthy, M. J. (1998). *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.
McCarthy, M. J. (2003). Talking back: "small" interactional response tokens in everyday conversation. *Research on Language in Social Interaction, 36*(1), 33–63.
McEnery, T., & Wilson, A. (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.
Müller, S. (2005). *Discourse markers in native and non-native English discourse*. Amsterdam: John Benjamins.
O'Keeffe, A., & Walsh, S. (2012). Applying corpus linguistics and conversation analysis in the investigation of small group teaching in higher education. *Corpus Linguistics and Linguistic Theory, 8*(1), 159–181.
O'Keeffe, A., Clancy, B., & Adolphs, B. (2011). *Introducing pragmatics in use*. London/New York: Routledge.
Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analysing data in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 351–383). Thousand Oaks: Sage.
Othman, Z. (2010). The use of *okay, right* and *yeah* in academic lectures by native speaker lecturers: Their "anticipated" and "real" meanings. *Discourse Studies, 12*(5), 665–681.
Polat, B. (2011). Investigating acquisition of discourse markers through a developmental learner corpus. *Journal of Pragmatics, 43*(15), 3745–3756.
Romero-Trillo, J. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics, 34*, 769–784.
Romero-Trillo, J. (2012). Pragmatic markers. In *Encyclopedia of applied linguistics* (pp. 4522–4528). Oxford: Blackwell-Wiley.
Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn taking for conversation. *Language, 50*, 696–735.
Schegloff, E. A. (1997). Practices and actions: Boundary cases of other-initiated repair. *Discourse Processes, 23*(3), 499–545.
Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.
Schiffrin, D. (2001). Discourse markers: Language, meaning and context. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 54–75). Oxford: Blackwell.
Schleef, E. (2008). Gender and academic discourse: global restrictions and local possibilities. *Language in Society, 37*, 515–538.
Schourup, L. (1999). Discourse markers. *Lingua, 107*, 227–265.
Scott, M. (2008). *WordSmith tools* (Version 5). Liverpool: Lexical Analysis Software.
Seedhouse, P. (2004). *The interactional architecture of the language classroom: A conversation analysis perspective*. Oxford: Blackwell.
Sinclair, J. (2004). Intuition and annotation: The discussion continues. In K. Aijmer & B. Altenberg (Eds.), *Advances in corpus linguistics* (pp. 39–59). Amsterdam: Rodopi.
Tay, D. (2011). Discourse markers as metaphor signalling devices in psychotherapeutic talk. *Language and Communication, 31*(4), 310–317.
Ten Have, P. (2007). *Doing conversation analysis: A practical guide*. London: Sage.
Trester, A. M. (2009). Discourse marker 'oh' as a means for realizing the identity potential of constructed dialogue in interaction. *Journal of Sociolinguistics, 13*(2), 147–168.
Walsh, S. (2006). *Investigating classroom discourse*. London/New York: Routledge.
Walsh, S. (2011). *Exploring classroom discourse: Language in action*. London: Routledge.
Walsh, S., Morton, T., & O'Keeffe, A. (2011). Analyzing university spoken interaction: A CL/CA approach. *International Journal of Corpus Linguistics, 16*(3), 326–345.
Wooffitt, R. (2005). *Conversation analysis and discourse analysis: A comparative and critical introduction*. London: Sage.

Part IV
Reviews

Review of Götz, S. (2013) *Fluency in Native and Nonnative English Speech*. Amsterdam: John Benjamins

Phoebe Lin

Abstract This chapter reviews Götz's book, Fluency in Native and Nonnative English Speech, which models speech fluency using quantitative corpus linguistic methodologies.

Speech fluency has long been a key area of concern in English language teaching (ELT). For decades linguists and ELT experts have been searching for ways to enhance learners' second language (L2) speech fluency. Some key topics that have been explored so far include the characteristics and assessment of fluent speech, cognitive and psycholinguistic processes underlying fluent speech production and possible interventions that may improve learners' fluency (e.g. pre-task planning time, familiarity with communicative task-specific vocabulary items, short-term residence abroad). When it comes to an examination of factors that contribute to the perception of speech fluency, previous studies tend to address one or two factors at a time. Götz's *Fluency in Native and Nonnative English Speech*, however, demonstrates that it is possible to explore the effects of, and interactions between, multiple fluency-contributing factors in a single study using quantitative corpus linguistic methodologies. This is a new development for speech fluency research.

The book presents a two-part empirical study in eight chapters. Chapter 1 introduces the rationale, aims and scope of the book and outlines a model of speech fluency which forms the backbone of the empirical study. The model consists of three components, each of which is reviewed separately in a following chapter. Chapter 2 begins with the first component, productive fluency, which comprises temporal variables (e.g. speech rate, mean length of runs, pauses), formulaic

Reviewed by Dr. Phoebe Lin, Polytechnic University of Hong Kong: phoebe.ms.lin@gmail.com

P. Lin (✉)
Department of English, Polytechnic University of Hong Kong,
Hung Hom, Hong Kong
e-mail: phoebe.ms.lin@gmail.com

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2,
DOI 10.1007/978-3-319-06007-1_14, © Springer International Publishing Switzerland 2014

sequences and the use of repeats, filled pauses, discourse markers (e.g. *well*, *you know* and *like*) and small words (e.g. *sort of*, *kind of* and *quite*). Chapter 3 presents the second component, perceptive fluency, which subsumes accuracy (i.e. number of errors), idiomaticity (defined as "things that are not wrong but sound odd" (p. 149)), intonation, accent, pragmatic features (defined as a speaker's competence to speak appropriately in given communicative situations), lexical diversity and sentence structure (defined as knowledge of a variety of possible syntactic structures and their subtle semantic differences as well as knowledge of discourse structure features such as cohesive devices). Chapter 4 briefly discusses the final component, nonverbal fluency, which includes a variety of body movements, including posture, facial expressions, and vocal sounds.

The report on the empirical study begins with the study's aims and methods in Chapter 5. The study sets out to examine the areas of speech which give away advanced level learners of English as NNSs. To do so, it compares the speech of 50 advanced level German learners of English (NNSs), from the Louvain International Database of Spoken English Interlanguage (LINDSEI-GE, 86,186 words), with the speech of 50 native speakers (NSs), from the Louvain Corpus of Native English Conversations (LOCNEC, 118,564 words). Chapter 6 presents the first part of the study where productive fluency variables from the two corpora were analysed using three inferential statistical tests. Firstly, loglikelihood ratio statistic G^2 indicates that the NNSs significantly overused filled and unfilled pauses and significantly underused 3-grams and 4-grams, repeats discourse markers and small words. They also had slower speech rates and shorter mean length of run compared to the NSs. Secondly, linear regression analyses reveal the duration of time spent abroad as the only significant predictor of the NNSs' temporal fluency ($t = 2.869$, $p < 0.01$). Finally, cluster analyses isolate six groups of NSs and NNSs with distinct productive fluency features and outcomes. Two NS groups achieved high temporal fluency either through frequent use of 3-grams or frequent use of discourse markers, small words and repeats. One NNS group achieved high temporal fluency through frequent use of 3-grams combined with infrequent filled pauses and repeats. The remaining groups of average to low temporal frequency are characterised either by: (1) their frequent use of filled pauses and repeats combined with infrequent use of 3-grams, discourse markers and small words, (2) their failure to use fluency enhancement strategies other than 3-grams, or (3) their frequent use of repeats, discourse markers and small words combined with infrequent use of 3-grams.

Chapter 7 reports on part two of the study which focuses on perceptive fluency. The audio recordings of five chosen samples from the NNS corpus were presented to a group of 50 NS raters composed of linguists and non-linguists. The raters needed to rate each NNS speech sample in terms of overall fluency and all seven perceptive fluency variables described in Chapter 3. Surprisingly, Pearson's r shows that the only factors that contributed significantly to the judgement of perceived overall fluency were, in ascending order of magnitude of correlation, accuracy as perceived by raters ($r = 0.89$, $p < 0.05$), accent ($r = 0.92$, $p < 0.05$) and the use of pragmatic features ($r = 0.93$, $p < 0.05$). No statistically significant correlation was found between temporal fluency variables from part one of the study and NS raters' overall fluency

ratings. Furthermore, observed accuracy, as measured by the actual number of errors made per hundred words, seems to be unrelated to accuracy perceived by raters, because the NNS with the least number of errors received the lowest ratings for perceived accuracy and the one with the highest number of errors received very high ratings for perceived accuracy. There was no significant correlation between observed fluency and perceived overall fluency either ($r = 0.56$, n.s.).

The final chapter, Chapter 8, summarises the major findings of the study and offers suggestions for future research.

While many factors have been said to contribute to L2 speech fluency, the empirical study successfully shows how much each factor actually counts towards perceived fluency from a statistical point of view. While the book discusses the implications of the empirical findings for ELT, the unsaid implications of the book for fluency research, particularly the finding about the disconnection between temporal measures of fluency and perceived fluency, appear as influential as its pedagogical implications, if not more.

Right from the outset, Götz outlines a model of speech fluency which differentiates between variables that contribute to productive fluency, perceptive fluency and nonverbal fluency. At the beginning, the model does not appear convincing. First, the attribution of each variable to productive or perceptive fluency seems arbitrary to some extent. The literature review chapters (Chapters 2, 3, and 4) argued for the relevance of the variables to speech fluency in general, but they did not help to justify why each variable should be attributed specifically to productive or perceptive fluency per se. Second, there are too many interconnections between all the fluency variables as the literature review chapters also point out. For example, the number of pauses is connected to speech rate, the mean length of run and frequency of errors, and formulaic language is connected to the number of pauses, hesitation and mean length of run. Therefore, it seems questionable whether the fluency variables can fit neatly into the dichotomy between productive fluency and perceptive fluency. Third, the reduction of the complex phenomenon of formulaic language to the frequency of 3-grams and 4-grams use in the empirical study is debatable. Besides, there are questions about the robustness of the n-grams because the two corpora are rather small for the extraction of n-grams. With the word combinations being highly variable and the corpora so small, the margins for error in the extraction process are elevated (see Lin 2010, 2013). Finally, the conceptual differences between many variables in Götz's model (e.g. small words versus discourse markers, formulaic language versus idiomaticity, pragmatic features versus idiomaticity) seem unclear.

However, support for the dichotomy between productive fluency and perceptive fluency indeed comes from the results of the empirical study. The absence of a statistically significant correlation between productive fluency variables and the overall perceived fluency ratings in part two of the study, combined with the disconnection between perceived accuracy and actual error frequency counts, points clearly towards the need to separate the judgement of overall perceived fluency from the objectively measurable productive fluency variables (e.g. speech rate, mean length of run, number of filled and unfilled pauses, error frequency counts).

Regarding the disconnection between perceptive fluency and productive fluency, Götz blames NS raters for being "not capable of evaluating and assessing any of the productive fluencemes accurately" (p. 87, see also p. 147) [note: a fluenceme is a label for an abstract factor that contributes to speech fluency, see p. 8]. However, it seems that the discrepancy merely reflects the fact that there is no one-to-one linear relationship between the measurable qualities of fluency (i.e. the productive fluency variables) and perception of fluency; it is not a matter of the raters' ability. In the field of intonation research, for instance, researchers acknowledge the close connection between the auditory/perceptual categories and the acoustic/instrumental categories (i.e. intonation is related to pitch, rhythm to duration of sounds and pauses, and stress to loudness). But, more importantly, they also acknowledge the fact that a one-to-one linear relationship between the measurable qualities and the perception does not exist. The perception of stress, for instance, is the result of interactions between loudness, pitch and duration. Results from the cluster analysis of the NS data—that high temporal fluency is connected with frequent use of either 3-grams or discourse management strategies incorporating discourse markers, small words and repeats—provide the evidence for the way in which the productive fluency variables interact with each other to produce a perception of fluency.

The use of spoken corpora in speech fluency research is undoubtedly an exciting development. However, the challenges confronting this approach are noticeable from reading the book. One of the challenges is the lack of metadata about learners' general L2 proficiency. Götz's study aims to shed light on the areas in which advanced level German learners of English fall short of the standards of NSs. Therefore, the credibility of the findings depends on the availability of objective information about the general L2 proficiency of the NNSs (e.g. their IELTS or TOEFL scores). In the absence of such an important piece of information, we can only *assume* the NNSs' advanced proficiency based on the fact that they were English majors at a German university. The need for this assumption may have weakened the credibility of the study.

All in all, the book has many strengths and is an important read for researchers of speech fluency and spoken corpus. The arguments are very clearly presented and easy to follow. There is a critical awareness of the limitations of the empirical study throughout. Most importantly, the book delivers in terms of inspiration and impact. While there have been many corpus studies reporting on NNSs' overuse and underuse of linguistic features compared to NS norms, this book shows that we may go further to investigate the interactions between multiple contributing factors and calculate the relative weighting of each factor. This method of using regression analysis in corpus linguistics to determine interactions between and relative weightings of contributing factors of a complex construct may not be new – Wulff (2008) applied it to model NSs' intuitive idiomaticity judgement. However, it may be the first time that the method is applied in speech fluency research. The types of statistical analyses have also expanded in this study to enable more perspectives on the corpus data. In addition to regression analysis, the book has also included loglikelihood and cluster analysis. In terms of research findings, the disconnection between the measurable factors of productive fluency and the

perception of overall fluency potentially challenges the practicality of the emphasis on measuring temporal features (including speech rate, number of pauses, mean length of run, phonation/time ratio) seen in many studies that aim to shed light on the teaching and learning of L2 speech fluency. The results of the study clearly show that these temporal features do not contribute significantly to perceived overall fluency. According to the results, there should be more focus on pragmatic features, accent and perceived accuracy in future fluency research or ELT, because they contribute significantly to perceived fluency.

References

Lin, P. M. S. (2010). *The prosody of formulaic language*. Unpublished PhD dissertation, University of Nottingham, Nottingham.

Lin, P. M. S. (2013). The prosody of idiomatic expressions in the IBM/Lancaster Spoken English Corpus. *International Journal of Corpus Linguistics, 18*(3), 561–588.

Wulff, S. (2008). *Rethinking idiomaticity: A usage-based approach*. London: Continuum.

Review of Kretzschmar, W.A. Jr. *The Linguistics of Speech* (2009) Cambridge: Cambridge University Press

Paweł Szudarski

Abstract The volume 'The Linguistics of Speech' is a description of the author's own model of speech seen as a complex system, highlighting its main aspects: the continuum of linguistic behavior, extensive variation, the importance of regional/ social proximity and differential frequency in social groups.

'The Linguistics of Speech' (LoS) is a description of Kretzschmar's model of speech seen as a complex system. Chapter by chapter, the author describes the intricacies of the model and highlights its main aspects: the continuum of linguistic behavior, extensive variation in all features, the importance of regional/social proximity to shared linguistic production and differential frequency as a factor influencing linguistic production in social groups. Given the complexity of these issues, Kretzschmar welcomes the advent of modern digital research tools such as multi-million corpora and explains how they can be used by contemporary linguists to analyze variation in language use. The LoS is presented as an alternative approach that does not reject traditional linguistic accounts (Kretzschmar focuses mainly on de Saussure's work on linguistic structure). Rather, it accompanies them and consequently enriches our understanding of spoken language in use.

The book consists of eight chapters which step-by-step introduce the LoS model to the reader. Chapter 1, 'The contemporary marketplace of ideas about language', offers an overview of different theories of and approaches to the formal study of language. Kretzschmar reviews the work of Bloomfield (structuralism), Chomsky (generativism) and Labov (sociolinguistics) in order to position the LoS as the study

Reviewed by Dr. Paweł Szudarski, Kazimierz Wielki University: pawel.szudarski@op.pl

P. Szudarski (✉)
Department of English, Kazimierz Wielki University in Bydgoszcz, Chodkiewicza 30, 85-064 Bydgoszcz, Poland
e-mail: pawel.szudarski@op.pl

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2, DOI 10.1007/978-3-319-06007-1_15, © Springer International Publishing Switzerland 2014

of variation and language in use. The author emphasizes that, unlike other approaches, the LoS does not make any assumptions about rule systems found in homogenous speech communities. Instead, it embraces variation as revealed by naturally occurring data. The chapter also discusses the Ebonics controversy as an example of how the same linguistic notion can be understood differently in academic and popular discourse. The former embraces the notion of correctness as seen in the institutionalization of Standard English, while the latter refers to the sense of rightness and appropriateness within a specific social group.

Chapter 2 is devoted to Ferdinand de Saussure and his contributions to linguistics. Kretzschmar recounts de Saussure's classic distinction between langue, the linguistics of linguistic structure, and parole, the linguistics of speech. According to de Saussure, the former is a primary concern of the linguist since it is linguistic structure that gives unity to language. Importantly, de Saussure did not consider speech to be unimportant, for he believed that the study of language consisted of two parts: structure and speech. In fact, he acknowledged that historically 'speech always takes precedence' (1916/1986: 19) over structure. Yet his work on language focused only on structure as a well-defined entity, which was partly caused by the fact that de Saussure did not have appropriate tools that are needed to study speech. Given that contemporary linguists have a greater number of resources at their disposal (e.g., large corpora of spoken data), due attention should be paid to the LoS as it offers a new way of analyzing variation found in speech.

Chapter 3, 'Evidence from linguistic survey research: basic description', explains how large-scale surveys such as, for example, the Linguistic Atlas of the Middle and South Atlantic States (LAMSAS) help linguists identify regional and social differences in the use of speech. Kretzschmar emphasizes that patterns of variation cannot be fully understood by an individual speaker whose linguistic experience is not reliable. It is only when one looks at large amounts of data such as surveys or corpora that the full extent of variability can be realized. Interestingly, as many examples from the LAMSAS show, the speech data produce a regular pattern of use that takes the form of the A-curve, an asymptotic hyperbolic curve that has been found to characterize the distribution of words (e.g., Ferrer-i-Cancho and Sole 2001) and formulaic sequences (O'Donnell and Ellis 2009). According to Kretzschmar, this property of speech is of key importance to the LoS model.

Chapter 4, 'Statistical evidence from linguistic survey research', focuses on the usefulness of statistics in revealing relationships between the distribution of speech variants and social and geographical variables. As Kretzschmar explains, if the LoS is to fulfill its purposes, it needs statistical tests to arrive at reliable findings that will represent the linguistic behavior of specific speakers living in specific regions. Locality is stressed as an organizational element of language-in-use. Consequently, if the LoS model purports to offer a continuum of linguistic behavior, it needs to take geographical distance into account ('proximity matters' p. 130). Since the distribution of speech variants on this continuum is a complex network of variables that are likely to correlate with one another, the linguist should rely on statistics in order to ensure validity and coherence.

Chapter 5, 'Evidence from corpus linguistics', focuses on corpora, that is, large computerized collections of naturally occurring data. Frequency analyses carried out by corpus linguists demonstrate that language use is a patterned behavior and that the distribution of words is not random. In fact, one can observe different kinds of phraseological partnerships called collocations which speakers use to convey meaning and create discourse. This suggests that the word is not necessarily the best unit of linguistic analysis. As Stubbs aptly comments, "it is not the words that tell you the meaning of the phrase, but the phrase which tells you the meaning of the words (2001: 14). There is a large body of phraseological research (e.g., Sinclair 1991; Erman and Warren 2000) which demonstrates how language is governed by the idiom principle, that is, how language users rely on prefabricated multi-word chunks functioning as single units. Kretzschmar welcomes these findings from corpus linguistics and explains how data from corpora ('the dimension of words in texts') and surveys ('the dimension of words in geographical and social space') are two interconnected dimensions of the LoS model that need to be analyzed together. Since they supplement each other, they can provide a better understanding of the distribution of speech. This is an interesting proposal and shows the interdisciplinary character of Kretzschmar's approach to the study of speech.

Chapter 6 discusses speech as a complex system. Drawing closely from Mandelbrot's (1982) work on complex systems, Kretzschmar relies on this analogy in his study of speech and thereby joins other linguists who have used this framework for linguistic analyses (e.g., Larsen-Freeman and Cameron 2008; Dörnyei 2010; De Bot et al. 2007; Ellis 2011). Speech as a complex system is described with reference to the following features: it is open and dynamic (therefore never in the state of equilibrium), it consists of many components, it shows an emergent order (hence its non-random distribution) and is characterized by non-linearity and scaling (similarly to maps, the same patterning can be studied at different levels of observation). The author also emphasizes that the non-linear distribution of speech is stipulated by Zipf's Law (Kretzschmar calls it 'the frequency of frequencies') which shows an inverse relationship between the frequencies of words in a text and the rank of these frequencies (put in descending order). As Kretzschmar explains, 'if one multiplies the frequency times the rank, the result is a number that remains roughly constant for every word in the text' (p. 190). Importantly, the curve is similar to the one drawn from the type/token ratio, a measure of lexical diversity in which the number of all running words in a text is divided by the number of different words. According to Kretzschmar, the advantage of Zipf's Law is that it reduces the level of variance in data. Finally, the fact that speech is a complex system does not mean that it is chaotic. Rather, it is probabilistic which is exemplified by the A-curve distribution of data.

Chapter 7 provides an interdisciplinary overview of the process of speech perception which involves psychology, neuroscience ('the contemporary counterpart of de Saussure's physiology'), cognitive anthropology (linking cognitive processes in individuals and whole populations), schema theory (schemas understood as 'arrays of slots for characteristics out of which a pattern is generated'), perceptual dialectology and spatial perception. Using evidence from Preston's

(1989, 1999) work in the USA and Horvath and Horvath's (2001, 2003) findings from Australia, Kretzschmar discusses the role of locality ('the concept of place') in the study of speech variants found in dialects. As already mentioned, the LoS is probabilistic since no individual is able to fully represent the collective behavior of the whole group they come from, and, vice versa, generalizations at higher levels of scales should not be applied to lower levels. This warns against the risk of relativism in the description of speech perception.

The final chapter deals with speech models and their applications. It offers a formal model of the system of speech based on all the arguments presented in the previous chapters. As the model involves quite complex mathematical operations, Kretzschmar reiterates his claims in a more discursive summary. It shows speech as a complex system in which production and perception influence each other in an emergent and self-organizing manner. The author also highlights the role of both social/geographical and textual factors as important dimensions of his model. Finally, the LoS is also related to language change and public policy.

All in all, the LoS is a welcome contribution to both sociolinguistics and corpus linguistics. It offers a dynamic approach to the study of speech which draws from linguistic surveys, corpus linguistics, dialectology, statistics, anthropology and cognitive linguistics. Even though some fragments of the book might seem daunting to readers who are unfamiliar with the topic, Kretzschmar is successful at elucidating the complexity of his model. He is convincing in presenting speech as a complex system in which linguistic behavior is a continuum representing the non-linear distribution of variation as dependent on proximity, scaling and the logic of aggregation. Such an understanding of language avoids making a priori assumptions and acknowledges how variation changes across geographical and social space. The LoS model points to the role of locality (social geography) and perception in accounting for variation encountered in language use. Finally, throughout all the chapters, the LoS is depicted only as an alternative approach to language which does not negate other frameworks. It depicts linguistics, the formal study of language, as an open marketplace where linguists exchange ideas without imposing one's views on others. It is thus worth stressing that the LoS does not reject de Saussure's linguistics of linguistic structure. In Kretzschmar's view, it is a different way of 'doing' linguistics which makes use of resources that were unavailable before and thereby enriches our understanding of language as a whole.

References

De Bot, K., Verspoor, M., & Lowie, W. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism, Language and Cognition, 10*, 7–21.

De Saussure, F. (1916/1986). *Course in general linguistics* (Harris, R, Trans.). LaSalle: Open Court.

Dörnyei, Z. (2010). The relationship between language aptitude and language learning motivation: Individual differences from a dynamic systems perspective. In E. Macaro (Ed.), *Continuum companion to second language acquisition* (pp. 247–267). London: Continuum.

Ellis, N. (2011). The emergence of language as a complex adaptive system. In J. Simpson (Ed.), *Routledge handbook of applied linguistics* (pp. 654–667). New York: Routledge.
Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*, 29–62.
Ferrer-i-Cancho, R., & Sole, R. V. (2001). The small world of human language. *Proceedings of the Royal Society of London, Series B: Biological Sciences, 268*, 2261–2265.
Horvath, B., & Horvath, R. (2001). A multilocality study of a sound change in progress: The case of/l/vocalization in New Zealand and Australian English. *Language Variation and Change, 13*, 37–58.
Horvath, B., & Horvath, R. (2003). A closer look at the constraint hierarchy: Order, contrast and geographical scale. *Language Variation and Change, 15*, 143–170.
Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford: Oxford University Press.
Mandelbrot, B. (1982). *The fractal geometry of nature*. San Francisco: Freeman.
O'Donnell, M. B. & Ellis, N. (2009, July 20–23). *Measuring formulaic language in corpora from the perspective of language as a complex system.* Paper read at 5th Corpus Linguistics Conference, University of Liverpool, Liverpool.
Preston, D. (1989). *Perceptual dialectology*. Dordrecht: Foris.
Preston, D. (1999). *Handbook of perceptual dialectology*. Amsterdam: John Benjamins.
Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: OUP.

Review of Partington, A., Duguid, A. & Taylor, C. (2013) *Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies (CADS)*. Amsterdam: John Benjamins

Róisín Ní Mhocháin

Abstract The volume Patterns and Meanings in Discourse: theory and practice in corpus-assisted discourse studies strongly argues the case for using the novel methodological approach of Corpus-Assisted Discourse Studies to unveil non-obvious meanings in discourse.

I was eager to read this book for numerous reasons not least that I was about to teach an MA module on language analysis and I am also intending to use corpus based methods in my own PhD research. All thought provoking ideas are, therefore, welcome. The book itself is a substantial volume with 12 chapters and an introduction, each chapter has a different but connected focus.

The introduction sets the scene for the book which claims to presume no prior knowledge of corpora studies, Corpus Linguistics (CL), Corpus-Assisted Discourse Studies (CADS), Discourse Analysis (DA) or even discourse. This claim is supported well with clear, easily digestible accounts of each being provided, accounts which would be relatively easy to follow for a student or researcher with limited experience in discourse studies or analysis. These accounts are developed as each chapter progresses with the authors opening up more focused and complex ideas for discussion and/or argument (see p. 33 for example).

The first page of this book outlines the shape of each chapter and the intended audience for the book. All of the content is clearly presented, although parts of this information is somewhat misleading as, in particular, while it states that suggestions for further research are included at the end of *each* chapter this is not in fact the case. Although the majority do indeed include quite extensive options for

Reviewed by Roisín Ní Mhocháin, Bath Spa University: r.nimhochain@bathspa.ac.uk

R. Ní Mhocháin (✉)
TESOL, School of Education, Bath Spa University, Newton Park,
Newton St Loe, Bath, BA2 9BN, UK
e-mail: r.nimhochain@bathspa.ac.uk

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2,
DOI 10.1007/978-3-319-06007-1_16, © Springer International Publishing Switzerland 2014

further study, not all chapters do. On the same page, the reader is introduced to the interesting selection of corpora used in the studies presented in the book and links them to the appropriate chapters. The corpora cover both spoken and written text although the latter appear to dominate.

Chapter 1 begins with two principles relating to the organisation of discourse, the idiom principle and open-choice; both are explained in terms of theory and the issues faced by researchers when working with idioms. The authors' analysis of *out of the corner of my/his eye,* using their SiBol 93 corpus, corroborated previous work carried out by Sinclair (2004) on the same idiom. The SiBol 93 (Siena-Bologna Modern Diachronic) corpus consists of all the articles published in three UK newspapers (*The Times/Sunday Times, The Telegraph/Sunday Telegraph* and the *Guardian*) in 1993.

Chapter 2 presents an interesting account of evaluation in discourse and highlights the importance of examining the context as part of the analysis. The authors' argument for the pervasiveness of evaluation in communication is strongly illustrated through the many extract examples and the supporting discussion. Connections are drawn between priming (from chapter 1) and evaluation, both of which are noted as being context-dependent phenomena.

Chapter 3 expands somewhat on chapter 2 adding the idea of control into the analysis and draws some interesting insights into control (the linguistic unit) and prosody relating to context. One item which can indicate the idea of control is *orchestrate* and we are shown examples of its use in both sport (positive use) and politics (negative use), where it is clear that context has a role to play in its interpretation. *Set in/sit through/undergo* and *budge* are some of the other items discussed as they have previously been looked at by other authors.

Chapters 4 and 5 both look at rhetoric in discourse (irony and metaphor specifically), and show how users of language challenge the idea of priming for rhetorical effect. Chapter 5 very helpfully begins by outlining some strategies for extracting conceptual mapping, showing that the novice researcher is never far from the authors' thoughts – a commendable feature of this book. Possible pitfalls identified by other authors are introduced and the central warning presented is 'that a corpus-assisted study can only tell you about the language in the corpus one is employing and therefore the composition of the corpus will necessarily affect the conclusions you reach'(p. 133). This is an important warning for novice researchers to heed. Many chapters include more than one case study and chapter 5 is no exception examining humour and metaphor.

Stylistics, commonly associated with literature, is examined in chapter 6. This was a particularly enjoyable chapter to read, not least due to the corpus chosen, the comic prose of P.G. Wodehouse! His work is chosen for examination because he is heralded as 'the greatest humorous novelist in the English language' (p. 168) and the challenge is to find evidence to show why this is so. In terms of novice researcher development, this chapter discusses why a CADS approach is beneficial for the study of stylistics in literature and leads the novice through the process starting with the puzzle to be examined (in this case a linguistic-stylistic study of Wodehouse's early work), developing the corpus for use, introducing other methodologies and techniques to employ (such as a close reading of the text) and then essentially identifying the step by step process for analysis. Links are very obviously drawn between this chapter and

those on irony and metaphor discussed earlier. Wodehouse's use of 'clashing' items relates also to priming as he pairs words in his text that would normally not collocate, my favourite example of those discussed is *deliberately love* and highlights the importance of close reading as such items would not appear in key word lists.

The authors advise readers about the difficulties of conducting cross-linguistic corpora studies in chapter 7, as words do not always translate according to their dictionary definition so the meaning and use is often lost in the ensuing translation, for example the metaphors discussed in chapter 5. Here they examine newspapers in both the UK and Italy and show that as well as issues with translation the features of the genre are not the same in both countries which must be taken into consideration during the planning and analysis stages of research. This is not the first time that context-specificity has been raised as we have seen it in previous chapters too.

Spoken corpora are the prime focus of chapters 8 and 9. The authors begin chapter 8 with a discussion of the cost and procedure involved in compiling a spoken corpus and also briefly discuss the linguistic differences between speech and writing. Institutional adversarial talk is the theme of the case studies in these chapters and while other authors (Harris 1991; Clayman and Heritage 2002; Hutchby 2005) have also investigated this area CADS is promoted for use in particular as it can 'reveal some of the non-obvious patterns of interaction as well as uncover the core items of routines, turns and speech acts' (p. 216).

The corpora used to investigate the theme of 'managing the message' in chapter 8 are from a selection of the White House briefings and the Hutton Inquiry and it is noted that all parties in both corpora would have had assistance in preparing for the question and answer format of both discourse types. The use of prefabricated chunks in speech is noted in the introduction to the chapter but the corpora include many longer and more complex prefabricated chunks than those employed in ordinary conversation possibly as the parties involved are 'professional discourse technicians' (p. 227). What is particularly interesting in this analysis is how forced priming works as evidenced by the White House briefings corpus; *Libya* was not mentioned in January of 2011 (p. 224 Figure 8.1) but in February the *Libyan government* along with lesser use of the *Libyan regime* and the *Quaddafi regime* appear, by March there is a distinct change in frequency of terms used and the *Quaddafi regime* is used almost twice as often as the *Libyan regime* and over three times more than the *Libyan government*. The most interesting finding is that the use of the *Libyan government* disappears from the journalists' speech with almost the same speed as it does from that of the White House Press Secretary. This case study, like all others carried out by the authors for this publication, employs a CADS approach.

CADS and (im)politeness, or more specifically the additional functions of negative politeness strategies, is the focus of chapter 9. This chapter begins with a summary of the relevant existing corpus-based research on the topic and identifies where CL can be useful but also indicates its limitations in this area. The benefits of using a CADS approach is again highlighted here showing the value of accessing the corpora in a number of ways: using CL software, actual close reading of the text and listening to the recordings. The final technique is important in terms of identification of *mock politeness* in particular. The expectation is that the book will not be read cover to cover

so in this chapter the subdivision of the Hutton Inquiry corpus is explained again and also the CADS methodology is reiterated. The conclusion reaffirms the potential of a CADS approach for a less researcher-dependent and therefore a more objective study of (im)politeness than they say has usually been the case.

In a move away from synchronic investigation, chapter 10 looks at language change in newspapers over a relatively recent period of time: presenting a modern diachronic study. As is the norm for this book, it begins with a look at what has been done to date and what this type of study entails along with the customary warnings for new researchers. Here the word of caution is in relation to patterns in the data and that different patterns will reveal themselves to different researchers which is in part influenced by our own individual primings. One of the initial findings of the research the authors conducted for this chapter shows that the formality of language used in newspapers is changing and less formal structures are now being employed. This is in addition to the unsurprising change in the use of proper names over the period (the years chosen for the corpora were 1993/2005/2010), so *Bush* replaces *Clinton* and *Blair* replaces *Major* with *Obama, Cameron* and *Clegg* appearing in the 2010 corpus. The issue of forced priming comes back into discussion again as the findings show that different newspapers were using the same quotations or phrases for certain parts of the newspaper which increased the number of times items appeared in the corpus and may be related to the use of promotional material in journalism.

A modern diachronic CADS approach is the focus of chapter 11 too however rather than language and discourse practices the focus here is on the change in attitudes to political, social and cultural issues. Two case studies are discussed one focusing on *antisemitism* (with no hyphen – the reason for which is detailed in the chapter) and the other looking at how *boy/s* and *girl/s* are represented. With the now expected duty of care towards those reading the book to upskill in discourse study methodology, the chapter details the methods of the approach to these studies as much as it discusses the results. As with the rest of the book, this chapter is well presented overall, however, occasionally the addition of some tables or graphs to illustrate the vast quantities of data put forward in the text (e.g. p. 298) would have been particularly helpful in assisting the reader in assimilating it.

The final chapter is a comprehensive discussion drawing together the various themes that run throughout the book. This includes further discussion on lexical priming which I have also noted as being a recurring theme. It also uses the studies in the book to show the eclecticism of CADS and admits that it is not an approach which is free from subjectivity. However, the researcher has no influence on the work of the software which will produce data faithful to the corpora used, regardless of the researcher's personal desires for the findings to support a hypothesis.

The strengths of the book are many. The recurrence of drawing together findings from previous analyses which relate to the focus of each chapter is one such and adds support to the authors' discussion of the analysis type being considered in each chapter. Added to this is the immense quantity, and quality, of the examples taken from the corpora to illustrate the points made, which can only be an asset for the inexperienced analyst in their quest for understanding. The layout of the book is

clear although the footnotes, while informative, are occasionally invasive and distract from the reader's train of thought (p. 49/50 for example). These longer foot notes could feasibly have been added to the end of each chapter.

The contents guide in itself is helpful in identifying the areas of interest for the analyses in each chapter and the type of analysis carried out in each case. The analysis types are also varied with cross linguistic analysis, comparison analysis and corpus-assisted stylistics to name a few. The appendix at the end of the book has very helpful information regarding corpus software, publicly available corpora and a selection of sources for downloading data for use in corpus research, data which were indeed utilised by the authors in their studies.

One of the many strengths of the book is that, although going from cover to cover makes for an interesting and coherent read, it is also entirely possible to simply dip in and out of the chapters relevant to your own study and absorb the many insights touched on in the discussion within.

Having read the book in detail and in its entirety, through the eyes of a tutor, a researcher and a scholar I can honestly say it was a very informative, often entertaining read. I have already recommended this book to my students as with guidance there are certainly many chapters that would be relevant and easily accessible to them. The book strongly argues the case for using CADS to help shed a light on *non-obvious meanings* (p. 11) in discourse and it can be said that it succeeds in shedding light on the many benefits of both CL and CADS techniques for the inexperienced researcher or student of linguistics.

References

Clayman, S., & Heritage, J. (2002). *The news interview: Journalists and public figures on the air*. Cambridge: Cambridge University Press.

Harris, S. (1991). Evasive action: How politicians respond to questions in political interviews. In P. Scannell (Ed.), *Broadcast talk* (pp. 76–99). London: Sage.

Hutchby, I. (2005). *Media talk: Conversation analysis and the study of broadcasting*. London: McGraw-Hill International.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge.

Author Index

A
Adelman, A.S., 2, 35–53
Adolphs, S., 272, 294, 297
Ahern, A., 8, 17, 22–24
Aijmer, K., 36, 214–216, 221, 222, 224, 225, 228, 233, 241–243, 249, 254, 292, 294, 297, 298, 310
Allan, K., 57, 63
Allen, G., 66, 152, 161
Altenberg, B., 36, 37
Álvarez, A., 241
Amador Moreno, C.P., 294
Amberber, M., 61, 73
Ameka, F.K., 58, 61
Amenós-Pons, J., 8
Amossy, R., 184
André, J., 118, 280
Angouri, J., 180
Antón Mellón, J., 197
Arche, M.J., 244
Aristar-Dry, H., 25
Arminen, I., 310
Arthur, M.B., 184
Artstein, R., 22
Asención-Delaney, Y., 38
Asher, N., 25
Aston, G., 38
Atwood, M., 115
Austin, J.L., 264, 267
Averill, J.R., 183
Ayala Torales, R.L., 162

B
Baayen, R.H., 12, 40, 44
Bach, K., 267
Bachman, L., 241
Baider, F., 3, 179–206
Bain, A., 25
Bainton, R., 162
Baker, M., 11
Bańko, M., 133
Banks, D., 187
Barcelona, A., 80, 91, 184
Bardovi-Harlig, K., 240, 241
Barker, M., 205
Barlow, M., 9
Barrett, L.F., 126
Bates, D.M., 45
Bazeley, P., 297
Beach,W., 302, 309
Bednarek, M., 135, 187
Beer, J.S., 125
Beers Fägersten, K., 57
Bell, D., 151
Benhabib, S., 183
Berkley, S., 232
Berlin, B., 102
Bernardini, S., 38
Besemeres, M., 74
Besnier, N., 206
Betz, H.D., 162
Biber, D., 219, 230
Bibok, K., 14
Bikowski, D., 262
Billig, M., 180, 183, 185, 187, 199
Bird, M., 155, 163, 166, 168, 172
Birner, B., 222
Bissonnette, C., 188
Blake, R.J., 262
Blakemore, D., 8, 241, 292
Blanco Gómez, M.L., 277
Blank, R.H., 160, 163
Blaye, A., 262, 263

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2, DOI 10.1007/978-3-319-06007-1, © Springer International Publishing Switzerland 2014

Blum-Kulka, S., 241
Blumczyński, P., 151
Blumenthal, P., 187
Boas, H.C., 9
Boggs, J., 262
Bolker, B., 45
Bosque, I., 242, 243, 249, 252
Boucher, J., 16, 17
Bouton, L., 253
Breen, M.P., 308
Breheny, R., 14
Breinin, E., 184
Brinton, L.J., 214, 299
Briz, A., 242
Bromhead, H., 61
Brown, B., 294, 297
Buchanan, S., 149–177
Bunce, C., 36
Burke, K., 185
Burridge, K., 57, 63
Buss, A.H., 126
Buswell, B.N., 126
Butler, J., 183
Buysse, L., 3, 213–235, 242

C

Callies, M., 215
Cambon, E., 198, 205
Cameron, L., 325
Campanelli, P., 294
Campillos Llanos, L., 3, 239–254
Campos, H., 243
Canale, M., 241
Candel Mora, M.A., 244
Caracelli, V.J., 298
Carletta, J., 22
Carretero, M., 4, 261–288
Carrió Pastor, M.L., 244
Carruthers, P., 16, 17
Carston, R., 8
Carter, R., 88, 242, 243, 253, 291, 292, 294, 299, 301, 310
Cartoni, B., 19
Casado Velarde, M., 242
Castle, J.M., 126
Cavalla, C., 187
Celce-Murcia, M., 241
Chafe, W., 41
Chambers, A., 294
Channell, J., 214, 230, 233
Charteris-Black, J., 185
Chenail, R. J., 309
Cheshire, J., 214, 221
Cheung, J.T., 187, 200
Chilton, P., 185, 186
Choi, I., 127
Cislaru, G., 67
Clancy, B., 188
Clancy, P.M., 41
Clarke, J., 104, 105, 113, 116–119
Clas, A., 61
Clayman, S., 331
Cole, M., 262
Collentine, J., 38
Collins, R., 73, 74
Comrie, B., 20
Conger, J.A., 184
Connor, U., 36
Conrad, S., 219, 230
Constantinou, M., 3, 179–206
Coon, H.M., 126, 127
Cornell, D., 183
Cosme, C., 36
Crawford, P., 294, 297
Crossley, S.A., 38
Culpeper, J., 57
Cummings, J.L., 57
Cummins, C., 9
Cutting, J., 214–216, 220, 224, 228, 233

D

Dagneaux, E., 244
Dailey-O’Cain, J., 293
Danino, N., 262
David, A., 244
Davidson, D.J., 40, 44
Davidson, R.J., 130
Davies, P., 197
De Bot, K., 325
De Cock, S., 215, 216, 219, 231
De Fina, A., 293
De la Fuente, M.J., 243
De Saussure, F., 324
De Saussure, L., 8, 17, 22
Degand, L., 241–243
Deignan, A., 80, 86, 88, 91, 93, 94
Deleuze, G., 206
Demonte, V., 242, 243, 249, 252
Denis, D., 214
Deshors, S.C., 37–40, 50, 52
Destruel, E., 25
Díaz Rodríguez, L., 253
Diener, E., 127
Diener, M., 127

Díez-Bedmar, M.B., 36
Díez Domínguez, P., 242
Dillenbourg, P., 262, 263
Dines, E.R., 214
Dirven, R., 20
Domenech Val, A., 243
Donald, M., 74
Dörnyei, Z., 296, 297, 325
Doğruöz, A.S., 51
Drouin, P., 188
Du Bois, J.W., 42
Dubois, S., 214
Duguid, A., 329–333
Dziwirek, K., 133, 135

E
Eatwell, R., 197
Edgeworth, R.J., 104, 105
Eggins, S., 295
Elder, C., 294
Ellis, N., 324, 325
Ellsworth, P.C., 130
Erman, B., 325
Escandell-Vidal, V., 8, 240
Evert, S., 219
Evison, J., 224

F
Fantauzzo, C.A., 126
Faraway, J.J., 44
Fassnacht, C., 302
Fernandez, J., 215, 216
Fernández Leborans, M.J., 252
Fernández López, S., 242
Ferrer-i-Cancho, R., 324
Fillmore, C., 124
Finegan, E., 219, 230
Fischer, K.W., 126, 128, 129, 144
Flowerdew, J., 294
Fodor, J.A., 28
Fontaine, J.R.J., 124
Foucault, M., 206
Francis, G., 11
Frank-Job, B., 292
Frank, R., 193
Fraser, B., 241, 291, 292, 299
Fraser, N., 183
Frijda, N.H., 182
Frye, M., 183
Fuller, J.M., 242
Fung, L., 242, 243, 292, 294, 299, 301, 310

G
Galindo Merino, M., 241
Garfinkel, H., 308
Garrido, A.M., 242
Garrison, D.H., 105, 109–111, 113, 117
Gass, S.M., 243, 253
Gaudin-Bordes, L., 184
Gaviño, V., 249
Gazdar, G., 8
Genung, J.F., 25
Gibbs, R.W., 10, 14, 15, 17, 28, 79
Gijsberts, M., 199
Gilbert, M.A., 124, 125
Gilquin, G., 10, 11, 15, 16, 216, 219, 242, 244
Girard, G., 187
Givón, T., 266, 268
Gladstone, W.E., 119
Goddard, C., 2, 55–75
Goffman, E., 145, 301
Goldsmith, H., 130
Golombek, P., 294
Goodwin, J., 183
Górski, R., 133
Götz, S., 4, 317–321
Gozalo Gómez, P., 3, 239–254
Graham, S.L., 57
Graham, W.F., 298
Granger, S., 13, 36–38, 243, 244
Granic, I., 182
Greene, J.C., 298
Grevisse, M., 20
Grice, H.P., 8, 14
Gries, S.T., 2, 10–12, 15, 16, 35–53, 82
Grisot, C., 1, 7–29
Grivaz, C., 13
Groenendyk, E., 182
Grossman, F., 187
Guirao, J.M., 246
Gutierrez, G., 163

H
Hainsworth, P., 197
Halliday, M.A.K., 91, 292
Hamilton, H.E., 66
Handl, S., 83
Hanks, P., 80, 91
Hardie, A., 80, 83, 92
Harnish, R.M., 267
Harrell, F.E.Jr., 47
Harris, M.J., 50
Harris, S., 331
Hart, D., 187

Hasan, R.H., 292
Hasselgård, H., 36, 37
Hasselgren, A., 215, 233
Haugh, M., 64
Haviland-Jones, J.M., 126
Hawkins, B., 193
Haynes, M.O., 126
Hellermann, J., 242
Henzl, V., 293
Heritage, J., 331
Hernández, T.A., 243, 253
Herring, S.C., 263, 278
Hidalgo Downing, L., 242
Hill, D., 61
Hinds, J., 41
Hinkel, E., 253
Hinrichs, E., 25
Hjelmslev, L., 101, 102
Hoey, M., 93
Hoffmann, S., 219
Hofstede, G., 126, 127, 145
Holmes, J., 57, 215, 220
Holmes, M., 182, 183
Hood, M., 262
Horace, F, 102
Horn, L.R., 8
Horvath, B., 326
Horvath, R., 326
House, J., 241, 253
House, R.J., 184
Hubbard, M., 105, 115
Hughes, G., 57
Hui, H.C., 127
Hundt, M., 36
Hung, J., 36–38, 215, 233
Hutchby, I., 331
Hymes, D., 241

I
Ieno, E.N., 44
Ilie, C., 193
Irwin, E., 119
Iwasaki, S., 43
Izadi, D., 215, 216
Izard, C.E., 126

J
Jackson, J., 182, 183
Jacobs, G.M., 308
Jacquey, E., 189
Jaeger, T.F., 40
James, C., 12
Janschewitz, K., 57
Jarvis, S., 38
Jasper, J.M., 183
Jay, T., 57
Jayez, J., 8, 25
Jefferson, G., 294
Jennings, E., 198
Johansson, M., 36, 38
Johansson, S., 12, 36, 37
Johnson, H., 103
Johnson, L.T., 155, 166
Johnson, M., 79–87, 93, 94, 184
Jucker, A.H., 100, 228, 291, 292
Jung, H.E., 294
Jung, J.Y., 241

K
Kahan, D., 183
Kahl, B., 155
Kamp, H., 20, 25
Kasper, G., 240, 241, 253, 305
Katsos, N., 9, 11
Kawaguchi, Y., 246
Kay, P., 102
Kecskes, I., 152, 182, 183, 206
Kellerman, H., 183
Keltner, D., 126
Kemmelmeier, M., 126, 127
Kemmer, S., 9
Kemper, T.D., 183
Kepser, S., 10
Kessler, G., 262
Kidman, A., 57, 61
Kirson, D., 126
Kitayama, S., 127, 182
Klein, D., 27
Knight, D., 302, 304, 309
Knowles, M., 83, 91
Kober, A.E., 104, 119
Koester, A., 228, 233
Kohl, D., 183, 186
Koike, D.A., 241, 253
Kopytowska, M., 124, 145, 146
Koselak, A., 189, 190
Kövecses, Z., 79, 80, 82, 83, 87, 91, 94, 124, 182
Kozłowska, M., 8, 25
Krauth-Gruber, S., 130
Kretzschmar, W.A.Jr, 4, 323–326
Kristeva, J., 199, 201, 205
Krzeszowski, T.P., 12, 37
Kumaravadivelu, B., 293
Kuno, S., 43
Kytö, M., 244

L
Labov, W., 20
Lafford, B.A., 244
Lakoff, G., 79–87, 93, 94
Langer, B., 241
Lardiere, D., 243
Larsen-Freeman, D., 325
Lascarides, A., 25
Lavid, J., 242
Lazarus, R., 182
Lee, D., 219
Lee-Goldman, R., 292
Leech, G.N., 12, 13, 298
Leeming, D.E., 262
Lehr, C., 128, 129
Leonetti, M., 8, 17, 18, 22, 24
Leow, R.P., 243
Lerot, J., 12, 13
Lesage, R., 188
Levenson, R.W., 126
Levin, S., 182
Levinson, S.C., 8
Levisen, C., 60
Lewandowska-Tomaszczyk, B., 3, 123–146
Lewis, M., 87, 126
Lewis, M.D., 182
Li, S., 2, 79–94
Liang, M., 93
Liao, Q., 249
Liao, S., 242, 248
Liebscher, G., 293
Lin, P.M.S., 317–321
Littlemore, J., 80, 83
Liu, B., 249
Ljung, M., 57
Llinares-García, A., 242, 294
Llobera, M., 241
Locher, M., 57
Longnecker, R.N., 153, 161, 162
Lorenzo-Dus, N., 241
Lowie, W., 325
Lozano, C., 244
Lubbers, M., 199
Lüdeling, A., 244
Luscher, J.-M., 25

M
Macaro, E., 325
Mackey, A., 243
MacMillan, M., 206
Maechler, M., 45
Mair, C., 61
Maíz-Arévalo, C., 4, 261–288
Mandelbrot, B., 325
Manning, C., 27
Marcus, G.E., 182
Marín Arrese, J.I., 4, 261–288
Markus, H.R., 127
Márquez-Reiter, R., 240
Marsden, E., 244
Martin, J.R., 266
Martin, R., 262
Martin, W., 12
Martín Zorraquino, M.A., 242, 243, 249, 252
Martínez-Caro, E., 242
Martínez, M.A., 4, 261–288
Maschler, Y., 292, 294, 301
Mashek, D.J., 144, 145
Masuda, T., 182
Matlock, T., 10
Matsuba, M.K., 208
Maxwell-Stuart, P.G., 118, 119
McAulay, L., 294
McCarthy, M., 242, 298
McEnery, T., 10, 11, 13, 57, 63, 80, 83, 92, 243, 298
McKinney, J.C., 308
Meara, P., 241
Meibauer, J., 9
Meier, R., 25
Mel'cuk, I.A., 61
Memmi, A., 205
Mendikoetxea, A., 244
Mettewie, L., 182, 185, 187
Mey, J.L., 2, 99–120
Meyer, T., 27
Mhochain, R., 329–333
Miglio, V.G., 50
Mio, J.S., 182, 184
Miracle, W.C., 249
Mitchell, R.L., 244
Moeschler, J., 7–29
Mohr, M., 57
Moise, J.F., 28
Montesa Peydró, S., 252
Montolío Durán, E., 242
Moon, R., 83, 91
Morelli, A., 182, 185, 187
Morency, P., 17
Moreno, A., 246, 294
Morris, G.H., 309
Morris, M.W., 127
Morton, T., 294, 297, 310
Mukherjee, J., 233
Mulder, J., 64

Müller, S., 215, 228, 233, 291–293, 299
Musgrave, S., 65
Musolff, A., 185, 187
Myles, F., 244

N
Nanos, M., 159, 163, 172
Neff van Aertselaer, J., 36
Németh, E.T., 14
Neu, J., 241, 253
Neumann, D.L., 262
Nicolle, S., 17
Niedenthal, P.M., 130
Nisbet, R.G.M., 105
Nisbett, R.E., 127
Nöel, S., 262
Nogueira da Silva, A., 243
Norenzayan, A.N., 127
Norrick, N.R., 57
Norris, P., 198
Novakova, I., 187
Noveck, I.A., 8, 9
Nussbaum, M., 183

O
O'Connor, C., 36
O'Donnell, M.B., 324
Ogarkova, A., 128, 129
O'Keeffe, A., 224, 242, 293, 295, 310
O'Malley, C., 262, 263
O'Neill, P.G., 101
Onfrey, M., 204
Ono, T., 41, 44
Onwuegbuzie, A.J., 297
O'Riordan, S., 294
Ortony, A., 79, 82, 86
Orts, M.A., 82
Osborne, J., 123–146
Östman, J.-O., 241–243
Othman, Z., 293
Overstreet, M., 223
Oyserman, D., 126, 127

P
Palacios Martínez, I.M., 221
Papaioannou, K., 209
Papp, S., 36
Paquot, M., 36, 37, 244
Partee, B., 25
Partington, A., 329–333
Parvaresh, V., 215, 216
Patterson, K., 93
Pavlenko, A., 74, 206
Pégard, C., 199
Peng, K., 127
Pérez de Ayala, S., 242
Pérez Hernández, L., 80, 88, 94
Perrineau, P., 197
Perry, B., 183
Pervin, L.A., 182
Petch-Tyson, S., 12, 36, 215, 233
Peters, P., 64
Peterson, E., 149
Pezik, P., 134
Piaget, J., 262
Piirainen, E., 2008
Placencia, M.E., 240
Plantin, C., 184
Plutchik, R., 183
Poesio, M., 22
Polat, B., 242, 292
Polleta, F., 183
Pons Bordería, S., 242, 252
Popescu-Belis, A., 28
Popper, M., 184
Porroche Ballesteros, M., 252
Portolés, José, 242, 243, 249, 252
Potvin, M., 182, 185, 187
Preston, D., 325
Prytz, Y.B., 219
Przepiórkowski, A., 133
Pullum, G.K., 102
Putnam, P.H., 126

R
Radden, G., 20
Ramírez Verdugo, D., 242
Rasekh, A.E., 215, 216
Rayias, M.F., 126
Rayson, P., 36, 51, 52
Reboul, A., 8, 25
Reese, R., 182
Reichenbach, H., 16, 17, 21, 23
Reiman, P., 262
Reips, U.D., 130
Reis, M., 10
Renandya, W.A., 308
Renouf, A., 294
Reyle, U., 25
Ric, F.,
Richards, J.C., 161
Riggio, R.E., 182
Robert, J., 262
Roberts, L., 25
Robins, R.W., 124, 126, 128, 144, 145, 187
Rodríguez-González, E., 253

Rodríguez Muñoz, F.J., 249
Roesch, E.B., 187
Rogatcheva, S., 36, 51, 52
Rojo, L.A., 82
Romero-Trillo, Jesús, 1–4, 233, 242, 291–294, 311
Rose, K., 240, 241, 253
Rosebury, B., 183
Rühlemann, C., 272
Ruiz de Mendoza Ibáñez, F.J., 80

S

Sacks, H., 294, 308
Sahota, O., 294, 297
Salaberry, M.R., 244
Salvan, G., 184
Sanchez Pardo, E., 242
Santana-Paixao, R., 50
Santos Río, L., 252
Savage, R., 183, 185, 186
Saveliev, A.A., 44
Scannell, P., 331
Scheeper, P., 199
Schegloff, E.A., 298
Scherer, K.R., 130, 182
Schiffrin, D., 242, 292, 294, 299, 300
Schleef, E., 292, 293, 310
Schmid, H.-J., 182
Schmied, J., 124, 145, 146
Schourup, L., 292, 299
Schulz, P., 185, 186
Schütze, C.T., 51
Schwartz, J., 126
Schwartz, S.H., 127
Scott, M., 216, 298, 302
Searle, J.R., 264, 267
Seedhouse, P., 307, 309
Sessarego, C., 241
Shamir, B., 184
Shapiro, M., 185
Shariff, A.F., 187, 200
Sharifian, F., 57, 58, 61, 67
Shaver, P., 126
Shibatani, M., 41
Shields J., 199, 206
Simon Gomez, M.A., 197
Simon-Vandenbergen, A.-M., 241–243, 249, 254, 292, 298
Simpson, J., 325
Sinclair, J.M., 87, 88, 187, 272, 298, 325, 330
Skandera, P., 61
Slade, Diane, 295
Smith, C.A., 182
Smith, C.S., 25
Smith, N., 8, 16
Sole, R.V., 324
Solin, A., 180, 181
Soriano, C., 124, 128, 129
Spada, H., 262
Sperber, D., 8, 9, 15–17, 23, 29, 124, 154, 241
Stalmaszczyk, P., 134
Stanger, C., 126
Stapleton, K., 57
Stefanowitsch, A., 82
Stein, D., 263, 278
Steinbach, M., 9
Steinmayer, O., 104
Stenström, A.B., 297, 310
Stern, D.H., 173
Stewart, D., 38
Sthioul, B., 25
Storch, N., 262
Stubbe, M., 57
Stuewig, J., 144, 147
Sullivan, M.W., 126
Swain, M., 241

T

Taavitsainen, I., 100
Tagliamonte, S.A., 214
Taguieff, P.A., 205
Takagaki, T., 246
Takagi, T., 41
Tangney, J.P., 144, 145
Tannen, D., 292
Tarouza, S., 294
Tashakkori, A., 297
Tavanger, M., 215, 216
Taylor, C., 329–333
Taylor, R.M., 241
Teddlie, C., 297
Ten Have, P., 297
Terraschke, A., 214–216, 220
Thomas, J., 241, 265
Thomas, R., 294
Thompson, S.A., 41, 44
Thurrell, S., 241
Tiryakian, E.A., 308
Tognini-Bonelli, E., 11
Tolstoy, L.N., 101
Tomasello, M., 9
Tono, Y., 11, 38, 243
Tracy, J.L., 124, 126, 128, 187, 200
Travis, C., 59

Triandis, H.C., 126, 127
Tumblety, J., 198–200
Tutin, A., 187
Tyne, H., 189

U
Uchida, S., 16, 17
Upton, T., 36

V
Van Dijk, T.A., 180–182, 187, 204
Van Lancker, D., 57
Vanhatalo, U., 61
Vaughan, E., 188
Ventola, E., 180, 181
Vergun, A., 242
Verschueren, J., 265
Verspoor, M., 325
Virgil, 104, 117, 118
Virtanen, T., 263, 278
Vogel, K., 36
Vygotsky, L.S., 262

W
Waletzky, J., 20
Walker, N., 44
Wallbott, H.G., 128, 129, 144
Walsh, S., 243, 292–295, 297, 300, 301, 305, 307, 309, 310
Ward, G., 222
Warren, B., 325
Waterman, A.S., 127
Wei, M., 242
Weigand, E., 265–267, 272
Weinrich, H., 20
Weiss, M., 126
Wheeler, E.M., 50
White, P.R.R., 266
Whitfield, F.J., 101
Wierzbicka, A., 56–61, 63, 67, 73, 74, 151, 156, 181, 182
Wildman, S., 200, 204
Wiley, N., 73
Wilkins, D.P., 58
Willems, D., 242
Williams, M.H., 197–200
Williams, R., 151, 152
Willner, A.R., 184
Wilson, A., 10, 13
Wilson, D., 8, 15–17, 23, 124
Wilson, P.A., 3, 123–146
Wooffitt, R., 298
Wulff, S., 38, 51, 320

X
Xiao, R., 11, 243

Y
Yang, S., 4, 291–311
Yao, J., 249
Yoshimi, D.R., 253
Yuldashev, A., 215, 216
Yule, G., 214, 223
Yumlu, K., 124, 145, 146
Yus, F., 267

Z
Zaima, S., 246
Zakay, E., 184
Zampolli, A., 124
Zhao, H., 242
Zienkowski, J., 241–243, 292, 298
Ziv, Y., 291, 292
Zuur, A.F., 44

Subject Index

A
Academic discourse, 292
ACE. *See* Australian Corpus of English (ACE)
Adjunctive, 214, 215, 217–225, 231, 232
Annotation, 12, 13, 16, 21–29, 40–44, 62, 64, 235, 243, 298–304
AntConc, 180, 188, 201
Appraisal, 129, 182, 266
Arabic, 171, 243, 270
Australian Corpus of English (ACE), 65, 66, 69
Australian English, 2, 55–75
Australian National Corpus (AusNC), 64, 65, 75

B
Bank of English (BoE), 80–90, 92, 94
Bible, 3, 106, 149–156, 163, 172, 174, 176, 177
Body, 11, 38, 59–61, 106, 109–110, 114, 191–195, 214, 227, 318, 325
British National Corpus (BNC), 85, 133–135, 137, 138

C
CADS. *See* Corpus-assisted discourse studies (CADS)
Cambridge and Nottingham Corpus of Discourse in English (CANCODE), 294
CDA. *See* Critical discourse analysis (CDA)
Celtic, 101, 153
Chinese, 42, 61, 129, 244, 248, 249, 254, 295, 296, 304, 305, 307, 310
Chi-square, 36, 37, 45, 46, 48, 52, 53, 134, 219, 231
Christian, 56, 57, 65, 68, 150, 152–154, 158–163, 167, 168, 173, 177, 204
CL. *See* Corpus linguistics (CL)
Classroom discourse, 293–297, 310
Cognitive, 8, 13, 17, 63, 67, 73, 79, 86, 88, 91–94, 124, 126, 133, 141, 181, 183–187, 240, 242, 263, 292, 294, 301, 302, 306, 307, 317
Collectivism, 3, 124, 126–128, 145, 146
Color blindness, 100, 103, 106, 119
Color relativity, 102
Color terms, 100, 102, 104–107, 112–114, 117–119
Common European Framework of Reference (CEFR), 240, 243, 244
Communicative competence, 241, 293
Computer mediated communication, 278
Conceptual Metaphor Theory, 79–94, 182
Concordance, 36, 39, 81, 82, 88, 90, 134, 136, 138, 140, 143, 216, 272, 302, 303
Connectives, 16, 26
Contempt, 3, 180, 183–186, 240, 242, 263, 292, 294, 301, 302, 307, 317, 325, 326
Context, 2, 8, 37, 56, 90, 100, 124, 181, 227, 240, 263, 292, 330
Contrastive, 2, 12, 16, 19–21, 36, 38, 43, 44, 47, 49, 50, 61, 123–146
Conversation, 2, 4, 8, 35–53, 57, 70, 71, 73, 74, 103, 125, 161, 215, 216, 235, 242, 246, 249, 251, 252, 270, 272, 276, 292–296, 301, 304, 307–310, 318, 331
Corpus-assisted discourse studies (CADS), 329–333
Corpus-based, 1–3, 12, 79–94, 124, 240, 242, 243, 253, 297, 331
Corpus linguistics (CL), 292, 294, 295, 297, 298, 303, 305, 307, 310, 331, 333
Correlation, 12, 13, 46, 50, 125, 132, 133, 144, 145, 304, 318, 319

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2,
DOI 10.1007/978-3-319-06007-1, © Springer International Publishing Switzerland 2014

Critical discourse analysis (CDA), 180, 181, 297
Cross-confessional, 177
Cross-cultural, 114, 126, 128, 151, 156, 180, 240, 292
Cross-linguistic, 2, 12, 18, 21, 26, 29, 36, 61, 124, 126, 156, 331, 333
Culture, 1, 2, 75, 124, 125, 128, 144, 145, 151, 186, 192, 205, 206

D
Danish, 101
Diachronic, 100, 330, 332
Dialectification, 124, 125
Discourse analysis, 180–185, 297
Discourse markers (DMs), 3–4, 26, 239–254, 291–311, 318–320
Disjunctive, 214, 215, 217, 218, 224–236
Dutch, 3, 36, 213–235, 244, 248

E
E-forums, 262, 271
Emoticons, 267, 272, 276, 279, 280, 282–285, 288
Emotion, 3, 73, 105, 123–146, 179–206, 266–268, 278, 282
Empirical pragmatics, 7–16, 28, 29
English, 2, 4, 7–29, 14, 36, 55–75, 80, 101, 124, 150, 191, 213–235, 242, 263, 293, 324, 330
English as a foreign language (EFL), 36, 215, 269, 292, 293, 295, 296, 304, 305, 307, 310, 311
English language teaching (ELT), 1, 319, 321
Eskimo, 102
Ethnopragmatics, 2, 57
Exbodiment, 125
Explication, 60–64, 66–70, 72
Expressives, 4, 60, 62, 68, 261–288
Externalization, 124, 125
Extreme-right, 180, 181, 184, 198, 199, 202, 204, 205
Eye-tracking, 9, 15

F
Finnish, 244
First language (FL), 240, 254
Flaccus, Q.H., 102
Fluency, 4, 249, 293, 298, 309, 317–321
Freedom, 3, 127, 149–177, 194, 201, 217
French, 3, 7–29, 36, 40, 105, 150, 151, 156, 157, 168, 169, 176, 180, 184, 198, 200–206, 215, 219, 231, 244, 248, 249

G
Galatians, 3, 149–177
General extenders, 3, 213–235
Generativism, 323
Geography, 107–109, 326
German, 3, 36, 100, 150, 151, 156, 157, 169, 176, 186, 215, 244, 248, 293, 318, 320
Givenness, 41, 45–50
Golden Dawn, 3, 179–206
Grammaticalization, 214, 300
GRAMPAL, 246
Greek, 104, 107, 116, 119, 151, 153, 155–157, 159–164, 166, 169, 173, 174, 176, 180, 188, 190–196, 205, 206
GRID, 3, 124, 126, 129–133, 135, 144
Guilt, 3, 124–126, 128–134, 136, 138–141, 144–146, 181, 186, 192, 193, 267, 273, 285

H
Hebrew, 151, 154, 158, 162, 164, 168–170, 173, 180, 186, 304
Historical pragmatics, 120
Hungarian, 244

I
Identity, 44, 124, 126, 129, 141, 145, 152, 157, 158, 160, 161, 164–168, 172, 173, 176, 180, 183, 198, 202, 204
Individualism, 3, 124, 126–128, 145, 146
Intercultural, 3, 103, 141, 156, 163–176, 270, 280, 284
Interjections, 2, 55–75, 267
Interlanguage, 38, 216, 240–242, 244, 254, 318
International Corpus of Learner English (ICLE), 36, 244
Inuit, 100
Italian, 29, 244, 248

J
Japanese, 2, 35–53, 101, 244, 248

K
Keyword, 3, 149–177
Korean, 42, 244, 270
KWIC, 188

L
Language acquisition, 292
Language pathologies neurolinguistics, 10

Latin, 2, 100, 102, 103, 153, 228
Learner corpus/corpora, 2, 35–53, 215–219, 221–224, 231–234, 243, 245
Learner English, 215, 230, 232, 233, 244
Lemma, 85, 87, 88
LOCNESS, 36
LookUp, 81
Louvain International Database of Spoken English Interlanguage (LINDSEI), 216, 219, 231, 244, 318

M

Machine translation (MT), 12, 16, 19, 27, 28, 150, 155, 156, 160, 161, 164–175
Metalexical awareness, 62–64, 68
Metaphor, 2, 79–94, 191, 330
Metonymy, 80
Monolingual, 11–13, 20, 21, 143
Multifactorial, 12, 38, 52
Multi-layered analytical approach, conversation analysis (CA), 291–311
Multilingual, 11–13

N

Narrative, 17, 20, 21, 23–27, 71, 74, 104, 142, 150, 159, 168, 243, 249, 300
National Corpus of Polish (NKJP), 133–137, 139
National Front, 3, 179–206
Native speaker (NS), 2, 3, 14, 22, 23, 35–53, 61, 215–219, 222–224, 231, 232, 234, 235, 239, 240, 244–254, 281, 293, 294, 318–320
Natural language processing (NLP), 9, 10, 13, 15, 16, 19, 27–29
Natural Semantic Metalanguage (NSM), 2, 57–64, 72, 73
Nature, 8, 15, 17, 18, 24, 26, 28, 38, 43–45, 49, 60, 66, 70, 80, 104–107, 114–117, 119, 125, 126, 167, 175, 214, 226, 230, 240, 249, 262, 276, 277, 280, 281, 284, 287, 288, 292, 310
Neo-Gricean, 8
Non-native speaker (NNS), 2, 4, 35–53, 240, 241, 245, 253, 254, 293, 294, 310, 320
Norwegian, 113, 215
NS. *See* Native speaker (NS)

O

Online collaborative learning, 263

P

Part-of-speech (POS), 135, 137, 243, 246, 298
Perception, 71, 83, 105, 125, 127, 168, 266, 280, 317, 320, 325, 326
Persian, 215
Persuasion, 184, 205
Phraseology, 100
Polish, 3, 61, 124, 126, 128, 130–137, 139–146, 244, 248, 270
Portuguese, 244, 246, 248
POS. *See* Part-of-speech (POS)
Pragmatic fossilization, 293
Pragmatic marker, 214, 215, 228, 229, 232, 233, 235, 241, 254
Pride, 3, 126, 146, 180, 185–187, 197–206
Procedural, 2, 7–29, 241, 307
Pro-concept, 17, 18
Psycholinguistics, 9, 15, 57, 84, 317

Q

Questionnaire, 2, 129, 241, 243

R

Regression, 2, 12, 38–40, 45–52, 318, 320
Relevance, 8, 15, 16, 21, 154, 241, 265, 292, 319
Rhetorical psychology, 180
Russian, 61, 270

S

Scripture, 150, 154, 158–160, 163, 168, 169, 173, 174
Second language acquisition (SLA), 292
Semantic molecules, 60, 68
Semantic primes, 58–60, 62
Semantics, 2, 8–10, 12–14, 16–19, 25, 37, 55–75, 88, 89, 94, 106, 124, 133, 135, 143, 156, 163, 174–176, 180–184, 186, 222, 242, 249, 254, 292, 299, 318
Shame, 3, 111, 118, 124–126, 128–138, 140, 142–146, 184
Siena-Bologna Modern Diachronic Corpus (SiBol), 330
Sky, 100, 116–117
Slavery, 3, 150, 152, 154–176
Slopeq, 134
Socialization, 124, 125, 240
Socio–cognitive pragmatics, 240
Sociolinguistics, 11, 50, 57, 64, 75, 323, 326

Spanish, 3, 36, 58, 59, 61, 150, 151, 156, 157, 175, 176, 204, 239–254, 270, 280
Spanish Learner Language Oral Corpus (SPLLOC), 244
Specificity, 21, 61, 180, 188, 202, 203, 331
Speech, 3, 4, 9, 16, 17, 21, 28, 56, 57, 59, 60, 64, 65, 67, 73, 100, 135, 137, 180, 181, 183–185, 187–203, 213–235, 241, 243, 245, 261–288, 298, 305, 309, 310, 317–321, 323–326, 331
Speech acts, 4, 9, 56, 100, 241, 245, 261–288, 331
Statistics, 2, 10, 12, 16, 22, 27, 36–38, 40, 44–46, 52, 130, 134, 180, 188, 201, 202, 217–219, 224, 231, 232, 234, 240, 247, 248, 253, 254, 264, 274, 288, 303, 307, 318–320, 324, 326
Structuralism, 323
Subjective, 10, 17, 38, 60, 63, 64, 73, 74, 129
Swearing, 57, 58, 61, 64, 67
Symbolism, 3, 185

T
Target language (TL), 3, 9, 12, 18, 19, 141, 231, 233, 234
TermoStat, 180, 188
Theoretical pragmatics, 1, 7–29
Time, 2, 9, 37, 56, 79–95, 100, 150, 187, 215, 241, 263, 303, 317, 325, 330
TL. *See* Target language (TL)
Transana, 302–304
Translation, 2, 3, 7–29, 141–143, 149–177, 191, 242, 331
Turkish, 244
Turn-taking, 294, 298, 300

V
Vague language, 215
Verb tense, 2, 8–10, 12, 16–22, 24, 25, 27–29

W
Wordsmith Tools, 140, 216, 298, 302, 303