

Georgia State University

ScholarWorks @ Georgia State University

Applied Linguistics and English as a Second
Language Dissertations

Department of Applied Linguistics and English
as a Second Language

Summer 8-11-2020

Seeking the unseen humanities macrostructures: The use of corpus- and genre-assisted research methodologies to analyze written norms in English and Spanish literary criticism articles

William Lake

Follow this and additional works at: https://scholarworks.gsu.edu/alesl_diss

Recommended Citation

Lake, William, "Seeking the unseen humanities macrostructures: The use of corpus- and genre-assisted research methodologies to analyze written norms in English and Spanish literary criticism articles." Dissertation, Georgia State University, 2020.
https://scholarworks.gsu.edu/alesl_diss/56

This Dissertation is brought to you for free and open access by the Department of Applied Linguistics and English as a Second Language at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Applied Linguistics and English as a Second Language Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

SEEKING THE UNSEEN HUMANITIES MACROSTRUCTURES:
THE USE OF CORPUS- AND GENRE-ASSISTED RESEARCH METHODOLOGIES TO
ANALYZE WRITTEN NORMS IN ENGLISH AND SPANISH LITERARY CRITICISM
ARTICLES

by

WILLIAM MICHAEL LAKE

Under the Direction of Viviana Cortes, PhD

ABSTRACT

Descriptive studies of general and discipline-specific academic writing genre conventions have paved the way for pedagogical materials that build real-world skills for novice academic writers. To name some better-known cases, breakthroughs have taken place in this regard in the fields of psychology, engineering, and chemistry. However, attested scholarship on rhetorical patterns in humanities writing, such as published literary criticism (hereafter “LC”) is less common. This dearth of research affects scholars of literature produced by Spanish-speakers who write in both English and Spanish. Many L1 Spanish user scholars must often publish their research in English, rather than Spanish, to maintain institutional employment. Postsecondary

Spanish majors in the U.S. must also demonstrate competence in literary criticism to gain credentials. To address the needs of these groups, the present study examines the potential of lexical bundles, qualitative content, and multidimensional analyses to help describe LC from a lexico-grammatical perspective. Such findings may facilitate an arrival at a comprehensive schematic of strategies used by expert-level literary scholars in Spanish and English. First, using multidimensional analysis, linguistic features characteristic of literary criticism writing are analyzed and interpreted in the context of prior multidimensional analyses to offer insight on ways in which the written norms of LC compare to those espoused in other genres previously analyzed. Next, the study examines the syntactic structures and functions of lexical bundles used in English and Spanish LC writing, with particular attention to quasi-equivalent and language-specific bundles. Finally, the study proposes a taxonomy of communicative strategies utilized by literary scholars in their arguments. Devised via qualitative content analysis, this taxonomy may extend the functional analysis of bundles in LC. These findings offer further insight into the macrostructures of literary criticism, as well as the sentence-level strategies that serve as building blocks for expert-level writing in the genre.

INDEX WORDS: Literary Criticism, Humanities, English for academic purposes, Spanish for academic purposes, Lexical bundles, Parallel corpora

SEEKING THE UNSEEN HUMANITIES MACROSTRUCTURES:
THE USE OF CORPUS- AND GENRE-ASSISTED RESEARCH METHODOLOGIES TO
ANALYZE WRITTEN NORMS IN ENGLISH AND SPANISH LITERARY CRITICISM
ARTICLES

by

WILLIAM MICHAEL LAKE

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2020

Copyright by
William Michael Lake
2020

SEEKING THE UNSEEN HUMANITIES MACROSTRUCTURES:
THE USE OF CORPUS- AND GENRE-ASSISTED RESEARCH METHODOLOGIES TO
ANALYZE WRITTEN NORMS IN ENGLISH AND SPANISH LITERARY CRITICISM
ARTICLES

by

WILLIAM MICHAEL LAKE

Committee Chair: Viviana Cortes

Committee: Diane Belcher

Stephanie Lindemann

William J. Nichols

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

August 2020

DEDICATION

In memoriam

Drs. Gracia Roldán (dec. 2017) & William Ormond Deaver Jr. (dec. 2019).

May this work continue the momentum you set forth toward humanity's growing power to reason with and through literature in all languages of the world.

ACKNOWLEDGEMENTS

First and foremost, much gratitude is to my advisor Viviana Cortes for galactic proportions of leadership and inspiration to take a reasoned approach toward endless data sources, resulting in a path of temperance toward vast questions. I am also indebted to the dissertation committee members Diane Belcher for highly meaningful discussions and for opening the door to studies of academic journal articles, Stephanie Lindemann for garnering interest in ideas that I never conceived of as possible, and William J. Nichols whose guidance for the past 9 years at GSU has allowed for opportunities to teach and research in the field.

Much credit is due to René Venegas, who graciously provided the encouragement and resources necessary to replicate his 2010 study. Additionally, many thanks to Bethany Gray for extensive support in fixtagging, Mark Davies, Douglas Biber, and Fernando Reati for insights on methodological questions, as well Jesse Egbert for facilitating the English-language tagging.

Much gratitude is also due to Victoria Rodrigo and YouJin Kim for their encouraging me to major in applied linguistics. I also thank everyone kind enough to offer an assistantship with no shortage of mentorship to boot: Óscar Moreno, Paula Garrett-Rucks, Carmen Hermann, Patricia Nolde, Héctor Fernández, Pete Swanson, Melissa Young, Deb Loden, Tim Jansa, Vicky Hanning, Bree Hicken, Jesse Rappole, Crystal Moody, and the Panther Athlete Support Services team, as well as those highly influential professors Ana Torres, Ellen Blossman, Mónica Díaz, Hugo Méndez, Annette Cash, Rudyard Alcocer, Leslie Marsh, María Elena Bermúdez, Raúl Llorente, Francis Howard, Gertrude Tinker Sachs, Jayoung Choi, Jeffrey Glover, Scott Crossley, John Murphy, Ute Römer, Sara Cushing, and Eric Friginal. Thank you as well to all graduate students who stopped by to say hello.

This research was also sponsored by the Honor Society of Phi Kappa Phi. Thank you to them as well as to the participants who graciously and patiently worked in the research paradigm. Finally, I thank my immediate family, with whom it would not have been in any fashion possible to continue with this project.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	V
LIST OF TABLES	XIV
LIST OF FIGURES	XVIII
LIST OF ABBREVIATIONS	XX
1	INTRODUCTION.....	1
1.1	Significance of the project	1
1.2	Research questions.....	4
1.3	Outline of the study	5
2	LITERATURE REVIEW	7
2.1	Literary criticism and academic writing.....	7
2.1.1	<i>Literary Criticism in the English and Spanish degree programs in the U.S.</i>	8
2.1.2	<i>Research on literary criticism norms in English</i>	9
2.1.3	<i>Research on literary criticism norms in Spanish.....</i>	13
2.1.4	<i>Descriptive Studies of Academic Writing</i>	15
2.1.5	<i>Multidimensional analysis</i>	17
2.1.6	<i>Subsequent MDA Studies.....</i>	18
2.2	Formulaic language antecedents of LBs	21
2.2.1	<i>Subsequent LB Studies in English</i>	23
2.2.2	<i>Studies on LBs in published Spanish writing.....</i>	24

2.3	Communicative Purposes	25
2.3.1	<i>Qualitative Content Analysis.....</i>	<i>27</i>
2.4	Summary: The motivation for the dissertation study.....	29
3	CORPORA AND METHODOLOGY.....	32
3.1	Corpora	32
3.1.1	<i>Register analysis</i>	<i>34</i>
3.2	Multidimensional Analysis (MDA) of linguistic features	36
3.2.1	<i>MDA in English.....</i>	<i>37</i>
3.2.2	<i>MDA in Spanish</i>	<i>49</i>
3.3	Lexical Bundle generation.....	56
3.3.1	<i>Structural Lexical Bundle analysis</i>	<i>57</i>
3.3.2	<i>Functional Lexical Bundle analysis</i>	<i>58</i>
3.4	Communicative Purpose identification	59
3.4.1	<i>Rater training</i>	<i>60</i>
3.4.2	<i>Passage selection</i>	<i>66</i>
3.4.3	<i>Lexical Bundle Communicative purpose analysis</i>	<i>70</i>
3.5	Conclusion.....	72
4	REGISTER ANALYSIS AND MULTIDIMENSIONAL ANALYSIS OF ENGLISH AND SPANISH CORPORA	74
4.1	Register Analysis of Literary Criticism in English and Spanish	74

4.1.1	<i>Participants</i>	76
4.1.2	<i>Textual layout and organization</i>	77
4.1.3	<i>Setting</i>	80
4.1.4	<i>Subject/Topic</i>	81
4.1.5	<i>Purpose</i>	81
4.1.6	<i>Nature of data or evidence</i>	82
4.1.7	<i>Methodology</i>	82
4.1.8	<i>Explicitness of research design</i>	82
4.2	RQ 1: Multidimensional Analysis of LCW in English and Spanish	84
4.2.1	<i>English MDA results</i>	85
4.2.2	<i>Dimension 1: Involved vs. Informational Production</i>	88
4.2.3	<i>Dimension 2: Narrative vs. Non-narrative Concerns</i>	95
4.2.4	<i>Dimension 3: Explicit vs. Situation-dependent reference</i>	102
4.2.5	<i>Dimension 4: Overt Expression of Persuasion</i>	106
4.2.6	<i>Dimension 5: Abstract vs. Non-abstract information</i>	109
4.2.7	<i>Dimension 6: On-line elaboration of information</i>	114
4.3	Spanish MDA results	118
4.3.1	<i>Dimension 1: Contextual and Interactive Focus</i>	122
4.3.2	<i>Dimension 2: Narrative Focus</i>	126
4.3.3	<i>Dimension 3: Commitment Focus</i>	132

4.3.4	<i>Dimension 4: Modalizing Focus</i>	137
4.3.5	<i>Dimension 5: Informational Focus</i>	140
4.4	Common threads in Dimension Scores in English and Spanish	144
4.5	Conclusion	145
5	LEXICAL BUNDLES IN ENGLISH AND SPANISH LITERARY CRITICISM	
	WRITING	148
5.1	RQ2: Lexical Bundles obtained	148
5.2	Structural analysis of lexical bundles	150
5.3	Functional analysis of lexical bundles	152
5.4	Equivalent analysis of lexical bundles	158
5.4.1	<i>Time-Event</i>	160
5.4.2	<i>Stance Epistemic Impersonal</i>	160
5.4.3	<i>Discourse Organisers: Topic Elaboration/Clarification</i>	162
5.4.4	<i>Topic Introduction / Focus</i>	163
5.4.5	<i>Referential: Intangible Framing Attributes</i>	164
5.4.6	<i>Summary of equivalent bundle analysis</i>	166
5.5	Language specific bundles	167
5.5.1	<i>Intangible Framing Attributes</i>	169
5.5.2	<i>Place Reference</i>	170
5.5.3	<i>Time-Event</i>	171

5.5.4	<i>Obligation-Directive</i>	172
5.5.5	<i>Structural only</i>	173
5.5.6	<i>Tangible Framing Attributes</i>	174
5.5.7	<i>Topic Elaboration / Clarification</i>	177
5.5.8	<i>Topic Identification / Focus</i>	178
5.5.9	<i>Stance</i>	180
5.6	Summary	182
6	COMMUNICATIVE PURPOSES AND LEXICAL BUNDLES	184
6.1	RQ3: Communicative Purposes Hierarchy	184
6.2	RQ4: Communicative purposes linked to lexical bundles	188
6.3	Equivalent Bundle Communicative Purpose Analysis	194
6.4	Agreement results	198
6.5	Saliency Results	201
6.6	The tactile experiencers and the elephant: why raters sometimes disagreed	203
6.7	Conclusion	204
7	CONCLUSION	207
7.1	Overview	207
7.2	Summary of Findings	208
7.2.1	<i>MDA-based register classification of literary criticism in English and Spanish</i>	208
7.2.2	<i>Function and structure of LBs in literary criticism</i>	209

7.2.3	<i>Communicative purposes of lexical bundles chosen</i>	210
7.3	Limitations	210
7.4	Pedagogical Implications	212
7.5	Suggestions for Future Research	219
7.6	Concluding Remarks	221
	REFERENCES	223
	APPENDICES	247
	Appendix A. English-language Training Document	247
	Appendix B. Spanish language Training Document	251
	Appendix C. Reference chart to aid participants' judgments	254
	Appendix D. Practice Worksheet – English	255
	Appendix E. Practice Worksheet – Spanish	257
	Appendix F. Four-word LBs from English corpus	260
	Appendix G. Five-word LBs from English corpus	263
	Appendix H. 6+-word LBs from English corpus	266
	Appendix I. Four-word LBs from Spanish corpus	269
	Appendix J. Five-word LBs from Spanish corpus	273
	Appendix K. 6+-word LBs from Spanish corpus	276
	Appendix L. All English LBs by functional category	277
	Appendix M. All Spanish LBs by functional category	278

Appendix N. Potentially English-specific LBs by functional category	280
Appendix O. Potentially Spanish-specific LBs by functional category	283
Appendix P. Heat map of Participants' categorizations: English	287
Appendix Q. Heat map of Raters' categorizations: Spanish	288
Appendix R. Informed Consent Form Georgia State University	289

LIST OF TABLES

Table 3.1. English corpus composition by country, journal, articles, and word count.....	33
Table 3.2. Spanish corpus composition by host country, journal, articles, and word count.....	34
Table 3.3. Situational characteristics framework (Gray, 2015)	35
Table 3.4. Description of Biber’s (1988) six dimensions and associated linguistic features	38
Table 3.5. Example of a tagged sentence.....	40
Table 3.6. Step 1: Obtaining frequency counts from which to calculate dimension scores	41
Table 3.7. Step 2: Calculating frequency counts normalized to occurrences per thousand words	42
Table 3.8. Step 3: Comparing normalized frequencies with means and SDs for linguistic features found in Biber (1988).....	43
Table 3.9. Summary of Dimension 2 calculation procedures for the abstract in Figure 3.1	43
Table 3.10. Comparison of the abstract tagged automatically by the Biber Tagger and tags fixed manually by the researcher.	47
Table 3.11. Spanish MDA features with examples, loadings used by Venegas (2010)	51
Table 3.12. Summary of Spanish DDS calculation process	54
Table 3.13. Functional categories of LBs with explanations of criteria and examples	58
Table 3.14. Communicative purposes hierarchy.....	59
<i>Table 3.15. Defining</i>	61
<i>Table 3.16. Describing</i>	61
<i>Table 3.17. Retelling</i>	62
Table 3.18. Explaining	63
Table 3.19. Evaluating	63
Table 3.20. Comparing	64

Table 3.21. Mean word count constituted by LBs per article (SD) in words and % of word count	67
Table 3.22. Word count, Article reference number, Tokens of LBs, and Research article section for texts chosen in the present study	67
Table 3.23. LBs by tokens found in the above passages (English).....	68
Table 3.24. LBs by tokens found in the above passages (Spanish)	69
Table 3.25. Agreement score calculation criteria	70
Table 4.1. Framework for describing the situational characteristics of academic journal registers	75
Table 4.2. Comparison of LCW articles and other disciplines (Gray, 2015) by page length.....	77
Table 4.3. English MDA results	85
Table 4.4. English MDA interfactor correlations	86
Table 4.5. Raw and % of frequency counts of linguistic features in Figure 4.2.....	90
Table 4.6. Grammatical features % of passage in Example 6	92
Table 4.7. Dimension 1 features subdivided as % of word count in Example 7	93
Table 4.8. Z-scores for normalized frequency counts of Dimension 3 features in English corpus	105
Table 4.9. Top 10 verb phrases triggering verb complement clauses.....	118
Table 4.10. Mean and (SD) of linguistic feature frequency counts, loadings, and products of linguistic features	120
Table 4.11. Literary Criticism Articles in comparison with other registers from Venegas (2010)	121
Table 4.12. Interfactor correlations for dimensional density scores in the Spanish LCAs.....	122

Table 5.1. Overlapping LB combination process	148
Table 5.2. Viability test for two four-word bundles and one five-word bundle	149
Table 5.3. Percentage of bundles divided by number of words from which they are comprised	149
Table 5.4. 4-word bundles attested in 5 studies and considered new in the present study	150
Table 5.5. Structural distribution of LBs in English and Spanish LCW.....	151
Table 5.6. Prepositions found in LBs by language	151
Table 5.7. Examples of English LB functional categories.....	153
Table 5.8. Examples of Spanish LB functional categories	154
Table 5.9. Distribution of LB functional category by language	157
Table 5.10. Equivalent LBs by functional category	158
Table 5.11. Language specific bundles by functional class.....	167
Table 5.12. Subgrouping of Spanish-specific tangible framing attributes bundles	174
Table 6.1 Communicative purposes hierarchy.....	185
Table 6.2. Lexical bundles used in sentences of each communicative purpose	185
Table 6.3. Equivalencies between bundle and communicative purpose across languages.....	194
Table 6.4. Partial agreement in communicative purpose judgments.	196
Table 6.5. Agreement calculation formula	198
Table 6.6. Mean agreement per English bundle.	199
Table 6.7. Mean agreement per Spanish LB.....	199
Table 6.8. Mean agreement and salience by functional bundle category in English.....	202
Table 6.9. Mean agreement and salience by functional bundle category in Spanish.	202
Table F.1. English 4-word bundles	261
Table G.1. English 6-word bundles	264

Table H.1. English 6+-word bundles	267
Table I.1. Spanish 4-word bundles	270
Table J.1. Spanish 5-word bundles	274
Table K.1. Spanish 6+-word bundles.....	276
Table L.1. All English LBs by functional category	277
Table M.1. All Spanish LBs by functional category	278
Table N.1. Potentially English-specific LBs by functional category.....	281
Table O.1. Potentially Spanish-specific bundles by functional category	284

LIST OF FIGURES

Figure 3.1. Abstract of The Aesthetics of Garbage in Philip K Dick’s Martian Time Slip.....	39
Figure 3.2. Mean Scores for Figure 3.1 and each genre in Biber (1988) on Dimension 2: Narrative Versus Non-Narrative Concerns.....	44
Figure 3.3. Spanish Abstract with dimensional features highlighted.	54
Figure 3.4. The abstract from Figure 3.3 classified according to its Dimension 2 features	56
Figure 3.5. Screenshot of Work Area	66
Figure 4.1. Geany’s search box and regular expressions detecting concentrations of positive- loading Dimension 1 features in proximity.	87
Figure 4.2. LCW articles in relation to other text varieties on Dimension 1 in English.....	89
Figure 4.3. LCW articles in relation to other text varieties on Dimension 2 in English.....	95
Figure 4.4. Tense tracking diagram of Example 10.....	99
Figure 4.5. LCW articles in relation to other text varieties on Dimension 3 in English.....	102
Figure 4.6.. LCW articles in relation to other text varieties on Dimension 4 in English.....	107
Figure 4.7. LCW articles in relation to other text varieties on Dimension 5 in English.....	110
Figure 4.8. LCW articles in relation to other text varieties on Dimension 6 in English.....	115
Figure 4.9. Spanish Dimension 1. Contextual and Interactive Focus.....	123
Figure 4.10. Spanish Dimension 2: Narrative Focus	126
Figure 4.11. Shifts in speaker time from Example 29.	128
Figure 4.12. Tense tracking diagram of verb tokens in Example 30.....	130
Figure 4.13. Spanish Dimension 3: Commitment focus arrangement of registers.	132
Figure 4.14. Tense-tracking diagram of Example 32.	134
Figure 4.15. Tense-tracking diagram of Example 33.	135

Figure 4.16. Spanish Dimension 4: Modalizing Focus.....	138
Figure 4.17. Spanish Dimension 5: Information focus.....	141
Figure C.1. Reference chart to air participants' judgments.	254
Figure P.1. Heat map of participants' categorizations of LB data in Spanish.....	287
Figure Q.1. Heat map of participants' categorizations of LB data in Spanish.	288

LIST OF ABBREVIATIONS

DDS	Dimension Density Score
LB	Lexical Bundle
LC	Literary Criticism
LCA	Literary Criticism Article
LCW	Literary Criticism Writing
MDA	Multidimensional Analysis
PMW	Per million words
PoS	Part of Speech
QCA	Qualitative Content Analysis

1 INTRODUCTION

1.1 Significance of the project

Recent decades have seen innumerable advancements in the teaching of discipline-specific writing as brought by the development of corpus-based descriptions of disciplinary norms. Three areas of study that have contributed to these understandings merit attention: genre-based studies of academic writing, data-based formulaic language research, and multidimensional analyses. In the case of genre-based scholarship on academic writing, we focus most on studies stemming from Swales' (1990) analysis of the communicative purposes present in research article introductions. Scholarship in this area has facilitated the teaching of academic writing to novices, leading to pedagogical materials for general academic (Swales, 1990; Bitchener, 2010; Bruce, 2008) and discipline-specific genre conventions (Atai & Falah, 2005; Stoller, Robinson & Jones, 2008; Kanoksilapatham, 2005; 2015). Qualitative Content Analysis (hereafter "QCA") (Mayring, 2015) is a mixed-methods approach to delineating linguistic features of texts into consistent, reliable categories. It has used qualitative, linguistic criteria to delineate characteristics of such under-analyzed text types as patient narratives (Engeli et al., 2016), children's interactions (Terveen, 2013), and textbook writing (Chen, 2019). As such, QCA may represent an area worth testing as a heuristic by which to examine the communicative purposes of rhetorical patterns in literary criticism.

Moving from the document and paragraph levels to the sentence level, lexical bundles (hereafter "LBs"), are clusters of three words or more that appear at high frequencies in a specific genre at an established frequency threshold per million words (Biber et al., 1999). From a formulaic language perspective, which informs research on LBs, it appears that knowledge and

use of collocations, or words which frequently co-occur with one another other (Firth, 1957), constitutes one element of proficient writing (Cortes, 2004). This assertion has received support in research on academic writing at the university level (Cortes, 2002; 2004; 2006; Shin, Cortes & Yoo, 2018; Shin & Kim, 2017; Rafieyan, 2018). Further, descriptive studies of lexical bundles in expert-level academic genres suggest that a wide variety of specific phrases may constitute required elements of learning to write in a given discipline (Biber et al., 1999; Chen & Baker, 2010; Cortes, 2004; Candarli & Jones, 2018; Gilmore & Millar, 2018; Jalali, 2014; Cortes, 2008).

Another technique which has made more evident the ways in which textual genres differ from one another is multidimensional analysis (hereafter “MDA”) (Biber, 1988). Research in this strand suggests that languages tend to diversify linguistic features by genre. For instance, spoken language tends to favor shorter sentences and significantly more second-person pronouns when compared to official documents in the same language (Biber, 1993; Biber & Hared, 1992; Kim & Biber, 1994, Besnier, 1988; Berber Sardinha et al., 2014; Biber et al., 2007). Further MDAs of academic writing have even evinced differences at the disciplinary and research paradigm level (Gray, 2015). Heightened access to scholarly publishing for writers of all backgrounds represents an area of great sociopolitical import for the teaching of academic writing in English (Belcher, 2007). As such, it is possible that comprehensive, empirical descriptions of written genre norms in academic disciplines could eventually be translated into teaching materials for academic writers.

While other European languages such as French and German have seen declines in U.S. university enrollment in the past 30 years, Spanish’s tendencies have remained relatively stable

in the U.S. university context. It also remains the most widely studied language at the postsecondary level in the U.S. However, while Spanish faculty have tended to favor the teaching of literature courses (Hertel & Dings, 2014), some students may hold more interest in practically oriented courses (Hertel & Dings, 2017). In the midst of this debate, Darhower (2014), having outlined a disconnect between theory and practice in the teaching of literary criticism writing (hereafter “LCW”) in Spanish programs, remarks that “applied linguists and literature faculty need to continue to work together” (p. 408) to overcome the increased scrutiny under which humanities programs find themselves (Modern Language Association, 2007; Francomano, 2012; Doyle, 2018; Nichols, 2019). Accordingly, research in the teaching of LCW at the university level warrants attention.

However, at present, there is little attestation of empirically founded techniques and their applications toward writing in the humanities (Tankó, 2017). Traditional approaches to the teaching of literary criticism writing, including classification of literary works into genres to facilitate discussion (Frye, 1957; Rosmarin, 1985), psychoanalytic interpretations (Hoffman, 1957) and a pedagogy based on the teaching of terms (Wilder, 2005) have yet to achieve groundbreaking success (Wilder, 2012). While an obvious solution would lie in partial replications of past genre-based studies of academic writing (cf. Swales, 1990; Moreno & Swales, 2018), one potential issue is the lack of standardization in formatting in humanities writing when compared to the prototypical research article structure preferred in other hard sciences (Biber & Gray, 2016). Tankó (2017), for instance, using rhetorical move analysis techniques (cf. Swales, 1990) on a corpus of literary criticism abstracts, argued that literary criticism abstracts adhered significantly less to norms of social science abstract writing. Yet one

may interpret this finding as a portrayal of literary criticism in terms of its deficits relative to social sciences rather than in terms of its intra-disciplinary assets. Regarding descriptive studies of LCW on its own terms, some minor success has been achieved with the teaching of topoi, or figuratively established places to which a researcher might frequently recur (Wilder, 2012). However, since lexico-grammatical features with said topoi are not yet attested in applied linguistics research to our knowledge, a reliable understanding of the role they play in the rhetorical strategies consistently used by literary critics may not be viable at the present.

To the author's knowledge, a descriptive study of LCW in English and Spanish informed by formulaic language, MDA analysis, and genre theory perspectives does not currently exist. Taking into account the vast benefits of various linguistic approaches to the charting out of norms in academic disciplinary writing from a variety of methods, the purpose of the proposed study is to devise a schematic for literary criticism articles (hereafter "LCAs") in top-tier English- and Spanish-language journals. Data for this schematic is derived via and interpreted in light of extant advancements on the use of corpus- and genre-assisted understandings of disciplinary writing. Specifically, we examine the following research questions:

1.2 Research questions

1. What can a multidimensional analysis in Spanish and English tell us about the correspondences between literary criticism and other genres?
2. What sorts of lexical bundles occur in literary criticism as a genre? What lexical bundles are equivalent in both languages and what lexical bundles appear language-specific?
3. What range of communicative purposes can be identified in a body of literary criticism articles? Is there any variation in this tendency in English and in Spanish?

4. Do any lexical bundles cluster more often than others toward the realization of specific communicative purposes?

1.3 Outline of the study

This dissertation's sections are ordered as follows. Chapter 2 contains a review of literature germane to the study of academic writing and literary criticism from corpus-driven and qualitative approaches, lexical bundles in academic writing, communicative purposes, and Qualitative Content Analysis.

Chapter 3 details the methods employed in this dissertation. It first describes how the parallel, comparable corpora were compiled and how their situational characteristics were assessed. It then describes the procedures by which the multidimensional, lexical bundle, and qualitative content analyses were conducted on data from both corpora.

Chapter 4 addresses Research Question 1 regarding the prevalence of linguistic features in LCW revealed via MDAs on the English- and Spanish-language corpora. Sub-sections in Chapter 4 show the results of the register analysis, noting situational characteristics in LCW that resemble and differ from writing in other disciplines, and the results of the English and Spanish MDAs, noting the lexicogrammatical profile revealed to be unique to LCW.

Chapter 5 examines Research Question 2 regarding lexical bundles in LCW. Sub-sections in Chapter 5 contain functional and structural analyses of these expressions, as well as a comparison of equivalent & quasi-equivalent bundles across languages as well as language-specific LBs.

Chapter 6 addresses Research Questions 3 and 4 regarding lexical bundles which cluster to communicative purposes and the range of communicative purposes realized in LCW. The subsections identify the range of communicative purposes detected in the corpus, analyze the results of the Qualitative Content Analysis conducted with 2 English- and 2 Spanish-speaking graduate students in literary criticism, and discusses a set of lexical bundles which appear to be linked to the realization of some specific communicative purposes.

Chapter 7, finally, contains a summary of the primary findings described in its preceding chapters. It also offers remarks on the dissertation's limitations, pedagogical applications, and recommendations for further research.

2 LITERATURE REVIEW

The present section contains a state-of-the-art review of the literature on academic writing research deemed relevant to the study of published academic literary criticism writing in Spanish and English. Specifically, Section 2.1 provides an operationalization of literary criticism writing, covers its centrality in Spanish degree programs in the United States, reviews research on literary criticism in Spanish and English, and discusses the Biberian and Swalesian approaches to studying academic writing in other disciplines which may enrich our understanding of LCW. In Section 2.2, relevant background literature on formulaic language and lexical bundles is summarized. Discussing the types of language functions achieved in a given document, Section 2.3 summarizes the construct of communicative purposes, literature to which the concept has been applied, as well as Qualitative Content Analysis (Mayring, 2014) which may represent one approach by which to study communicative purposes. Section 2.4 concludes the chapter by stating the motivation for the dissertation study. Taken together, the literature examined in this section builds a foundation for the study to be presented in this dissertation.

2.1 Literary criticism and academic writing

While social sciences are informed by positivist and postpositivist notions of identifying generalizable laws for social and mental phenomena (Miles et al., 2014), Kusch (2016) proposes that literature “offers us the human life in total — not reduced — with its inconsistent logic, morality, and identity on full display” (p. 1). Viera Echevarría (2014), however, argues that one does find a certain degree of structure and purpose in LCW:

“En literatura [...] el desarrollo se construye como el espacio donde se discute el análisis de la obra o proceso social y se busca reafirmar la tesis, usualmente mediante la presentación de ejemplos del texto literario analizado o mediante la presentación de citas de críticos literarios reconocidos”

[In literature, the development [of an argument] is constructed as the space in which one discusses the analysis of the text or social process and aims to defend an argument, usually via the presentation of examples from the literary text analyzed or via the presentation of quotes from renowned literary critics] (p. 323, translation mine).

Despite these assertions, there remain some unanswered questions as to how one, from start to finish, goes about producing an analysis of a literary work. To that end, in the present section, we review literature relevant to the teaching of literary criticism writing, identifying future directions for research therein.

2.1.1 Literary Criticism in the English and Spanish degree programs in the U.S.

Studying works of literature remains a fixture of foreign language degree programs in the United States (Hertel & Dings, 2014; 2017). L1 and L2 Spanish users at U.S. universities find themselves tasked with becoming proficient in writing within scientific and humanities genres (Modern Language Association, 2007; Francomano, 2012; Klee, 2015; Hertel & Dings, 2017). Proponents of maintaining ties between literature as part of foreign language degree programs argue that potential boons include repeated opportunities to acquire less-frequent vocabulary, insights into cultural perspectives more readily available outside the periphery of one’s native

culture (Nance, 2015), scaffolds for higher-order hypothetical discussion practice (Henning, 1993; Pearson et al., 2006), and heightened abilities to reason effectively within a given set of circumstances (Nichols, 2019). Nance (2015), conversely, contends that the extent of critical thinking or fluency gains may depend on the pedagogical methods used by literature professors.

For the purposes of this dissertation, it must also be stated that learning the genre-specific features of literary criticism is not solely a challenge for L2 Spanish users in Spanish degree programs at the graduate level. Two top-tier journals dedicated in part to Spanish-language works of literature, *Hispania* and the *Bulletin of Hispanic Studies*, accept submissions in both English and Spanish. However, many university administrators around the world regard a publication written in English as more prestigious than one written in a scholar's native language, if other than English (Swales, 1997; Belcher, 2007; Pérez-Llantada, 2015). As such, becoming flexible in literary criticism norms in both English and Spanish may be regarded as a crucial skillset for experts in the field of Spanish-language literary studies.

2.1.2 Research on literary criticism norms in English

Scholarship on works of literature has, historically, consisted of classifying a work as belonging to a given genre in order to open up further discussions about the work's importance (Rosmarin, 1985). Indeed, literary criticism may still be perceived by some as a discipline of little practical value that operates within a nebulous boundary between evidence and opinion (Small, 2015). In turn, universities have seen a decrease in funding for the teaching of literature in both English (Small, 2015) and Spanish (Nance, 2015). Some literary critics, on the other hand, maintain that training students to classify a work into a genre facilitates the creative process for aspiring artists, while also helping educators to teach cultural dimensions of literary

works to both their students and the public (Frye, 1957). For instance, using Ancient Greek syllogisms, it is possible to open cultural discussions of a novel by speculating on the genre to which it belongs. Rosmarin's (1985) categorization heuristic for genre classification, for example, could be applied to a novel familiar to U.S. American readers as follows: define "incendiary novels" as novels that are "seditious" (Shapiro, 2004) to the status quo; discuss Harriet Beecher Stowe's *Uncle Tom's Cabin* as an incendiary novel (Gill, 1996); finally, discuss *Uncle Tom's Cabin* as holding messages that were "seditious" to the status quo of the mid-19th Century U.S.

However, in more recent years, literary studies have also come to involve using a novel to approach a societal problem from an interdisciplinary approach (Tötösy de Zepetnek, 2007). This shift in the field is exemplified well in the following abstract excerpt from a recent article in the journal of Multiethnic Literature of the U.S. (MELUS).

"Sandra Cisneros's *Caramelo*, or, *Puro Cuento*: A Novel (2002) dramatizes the functions of travel and tourism for members of the Mexican and Chicana/o diaspora, particularly for second-generation Chicana protagonist and narrator, Lala Reyes. [...] Drawing from relevant strands in Chicana/o Studies, Indigenous Studies, Travel/Tourism Theory, Diaspora Studies, Postcolonial Theory, and Human Geography, I argue that Lala Reyes contends with multiple monolithic, colonially rooted views of what it means to be Mexican, US American, and Chicana, and that she ultimately forges a sense of self that refuses cultural binaries. In the process, she critiques and refutes expectations of cultural purity

that suggest migration and change over time necessitate a tainting or dilution of that identity.” (Szeghi, 2014, p. 162).

In the above example, the literary scholar explicitly puts forth an argument based not only on theoretical disciplines but also on Human Geography, which is a social science. To the author’s knowledge, no study has thus far paid attention to the kinds of language features that make up an interdisciplinary literary argument. Thus, it appears that more scholarship is needed.

At present, one finds little research that describes LCW from a linguistic perspective. A few exceptions to the general state of affairs include Wilder’s (2012) revised taxonomy (based on work by Fahnestock & Secor, 1991) of communicative strategies used in LCW. Wilder (2012) notes that it may behoove LCW instructors to emphasize teaching the components that make up a successful literary paper, rather than a general knowledge of literary genres and terms, an approach favored by past literary criticism textbooks (Blackwell & Larson, 2007). Conversely, Freedman (1993a) raises the following questions: “Is explicit teaching necessary? Is explicit teaching possible? If so, can it be useful? If so, when, at what stage in the evolution of a writer?” (p. 193) Negative answers to these questions are framed as the “Strong Hypothesis”, which argues that overt instruction of genre features is neither obligatory or helpful. Tentatively affirmative answers to Freedman’s questions form part of the “restricted hypothesis”, which proposes that lessons on certain features of a target genre may prove to be useful (Freedman, 1993a). Responding to Fahnestock (1993) in a later piece, Freedman (1993b) left open more possibility for future examinations of genres using more rigorous data collection and pedagogical approaches.

Concerning extant knowledge on genre conventions in LCW, MacDonald (1994) devised a taxonomy for the grammatical subjects of sentences written in research articles. Her results revealed that LCAs stand out for their higher-than-normal placement of observable persons and places in sentence subject position. Conversely, scientific articles favor the use of abstract entities such as statistical tests and past scholars as the agent of sentences. Thieme (2017) argues that the literary criticism article may sometimes contain a statement comparable to the methodology section in empirical articles, while acknowledging that when such a thing does occur, it is usually only a single noun phrase, such as a “close reading” conducted by one person. At the footnote level, Afros & Schryer (2009) offered a great deal of new insight on the kinds of promotional metadiscourse realized in literary studies’ footnotes, although not for the rest of the research articles in which they appear. At the paragraph level, *topoi*, the plural form of the Greek word *topos*, meaning place or site, with a metonymic relationship to a place established in the study, are one unit of analysis attested in later studies (Wilder, 2012; Viera Echevarría, 2014). *Topoi* may narrate plot details or establish the ubiquity of a phenomenon in support of one’s argument, frame of historical context, etc. (Fahnestock & Secor, 1991; Wilder, 2005; Wilder & Wolfe, 2009; Wilder, 2012). However, little is known about *topoi* from a linguistic perspective. For instance, comprehensive examinations of lexico-grammatical features associated with each *topos* in English and Spanish do not currently exist. Until further exploration of *topoi* is attempted, in the spirit of Darhower’s (2014) prescriptions for more collaboration between applied linguists and literary critics, a temporary solution may lie in the use of other qualitative research methods to devise more objective criteria by which to develop a schematic of the communicative purposes carried out in LCW.

Wilder (2012) conducted a study of teaching topoi to teachers, with largely successful results. One participant dissented, stating that systematically presenting conventions of literary criticism pedagogically would give students an “unfair advantage” (p. 185). From this professor’s point of view, he worked hard to obtain his status in the profession by acquiring disciplinary norms implicitly. While it is not known exactly to what degree other educators espouse such social Darwinism in their methods, the presence of any such individuals in the teaching profession warrants more research that could improve their students’ probability of success.

Wilder (2012) conducted a study of teaching topoi to teachers, with largely successful results. One participant dissented, stating that systematically presenting conventions of literary criticism pedagogically would give students an “unfair advantage” (p. 185). From this professor’s point of view, he worked hard to obtain his status in the profession by acquiring disciplinary norms implicitly. While it is not known exactly to what degree other educators espouse such social Darwinism in their methods, the presence of any such individuals in the teaching profession warrants more research that could improve their students’ probability of success.

2.1.3 Research on literary criticism norms in Spanish

As for LCW in Spanish, research tends to focus on the discipline’s sentence-level characteristics. For instance, Soto et al. (2005) found that humanities writing in Spanish had a noun to verb ratio of about 66%, whereas hard science articles sat at 78%. Beke and Bolívar (2009) found that philosophy writing was higher in modals of possibility compared to

psychology and education writing. Additionally, similar to findings in English academic writing (Biber & Gray, 2015), Soto & Zenteno (2004) observe that a feature that separates scientific writing from humanities writing is for events (e.g., when “Columbus discovered America in 1492” (p. 281, translation mine)) to be reformulated into what the authors call “pseudo-entity” noun phrases, for example, as “the discovery of America” (p. 281, translation mine) in scientific writing.

One highly cited text on the teaching of Spanish LCW is *Cómo se comenta un texto literario* (Lázaro & Correa, 1980), (*How to comment a literary text*, translation mine), which argues that a literary analysis can be crafted with knowledge of literary terms, a brief incorporation of the historical context, and the author’s biographical details. However, there is no mention in this text of how to include theoretical frameworks — a process seen as characteristic of the academic genre (Hoffman, 1957; Viera Echevarría, 2014).

One potential direction for short- and long-term improvements to LCW pedagogy may lie in Viera Echevarría’s (2014) analyses of transcripts from literary criticism presentations in Spanish. At the macrostructure level, Viera Echevarría (2014) found that literary criticism presentations may begin with a thesis statement followed by patterns of analyses of evidence and the use of additional sources to corroborate one’s position before drawing a conclusion. Although one characteristic of the mode of communication (Biber & Conrad, 2009) in a conference presentation is that it is spoken, rather than written, a high volume of literary criticism presentations at the academic conference from which she collected her data consisted of presenters reading written papers aloud. Thus, the degree to which her model for the rhetorical progression of literary criticism presentations holds value for written literary criticism articles

warrants an empirical examination. Moreover, few, if any, of the terms presented in a pedagogical approach to LCW based on teaching the definitions of literary terminology (Blackwell & Larson, 2007) appeared in Viera Echevarría's transcripts of the presentations. On the other hand, a keyword analysis revealed such terms as *memory*, *space*, *borders*, *bodies*, etc. These findings appear to support Biber's & Gray's (2010; 2016) assertion that specialist terminology in LCW is characterized by special operationalizations of everyday language, in contrast to the scientific neologisms prevalent in hard sciences.

2.1.4 Descriptive Studies of Academic Writing

In this section, we discuss two notable strands of research on describing the norms of academic writing: Swalesian rhetorical studies and Biberian Multidimensional Analysis. To begin, the theoretical underpinnings of Swalesian rhetorical studies (cf. Swales, 1990) emphasize an idea of genres as text varieties whose formal and functional features can be explained and taught after proper analysis by teacher-researchers. A cornerstone analytical technique in the Swalesian tradition is rhetorical moves analysis (Swales, 1990), in which segments of text are analyzed via top-down and bottom-up criteria to define their rhetorical purpose. From there, generalizable claims about the structure of text varieties are incorporated into descriptive and pedagogical interventions. Although not without its detractors from the beginning (Aull & Swales, 2015), the construct remains influential both in the scholarship and the pedagogy of academic English-language genres. After the initial success of the model's utilization to explain and predict characteristics of research article introductions (Swales, 1981), researchers branched out into examinations of other research article sections, including methods (Bruce, 2008), results

& discussions (Loan & Pramoolsook, 2015; Nodoushan & Khakbaz, 2011; Khany & Tazik, 2010; Atai & Falah, 2005; Ruiying & Allison, 2003), as well as more comprehensive studies spanning several sections (Kanoksilapatham, 2005; 2015; Cotos et al., 2015; Moreno & Swales, 2018). The model has also revealed that authors from primarily English-speaking countries may tend to criticize past studies more often than do scholars from Portuguese- and Spanish-speaking countries (Martín-Martín, 2005; Hirano, 2009; Sheldon, 2011). To the author's knowledge, the sole attempt at applying Swalesian methods to literary criticism abstracts (Tankó, 2017) was met with mixed results. The study found that many features, such as indicating the topic of study, recurred in nearly all abstracts in the sample. However, less consistency was found with other strategies, indicating that future studies may benefit from viewing LCW not in terms of its deficit-focused deviations from Swalesian models, but rather in terms of potential assets afforded by more flexible macrostructures. Indeed, if using a scheme such as Swales' in LCW is not suitable, it may be worth exploring other approaches to the study of communicative functions in this type of discourse.

As for empirical studies on academic writing in Spanish, sufficient knowledge has accumulated to produce instructional materials for L1 Spanish-speaking university students writing for science (Day, 2005) and humanities (Navarro, 2014) coursework. While there are also some Swalesian studies applied to Spanish-language research articles (e.g., Perales Escudero & Swales, 2011; Moreno & Swales, 2018; Martín-Martín, 2005; Otañi, 2005), other studies tend to focus on individual linguistic features in written academic disciplines, such as logical connectives (Mur Dueñas, 2009) and depersonalization features (Álvarez López, 2013). As such, a larger examination not only of individual linguistic features, but also the extent to which they

occur across a larger collection of texts and in conjunction with other communicative purposes is needed in Spanish. In the following section, we review Multidimensional Analyses literature, an approach that may address such a gap in the literature of humanities writing in English and Spanish.

2.1.5 Multidimensional analysis

Whereas Swalesian analyses are well-served by paying attention to phenomena at the sentence- or paragraph-levels with qualitative, quantitative, or mixed-methods approaches, multidimensional analysis (hereafter “MDA”) is a highly quantitative analytical technique that allows the researcher to arrive at the possibility of predictable generalizations about linguistic phenomena as they occur in an entire corpus. MDAs are based on exploratory factor analysis methodologies popularized in other social sciences (Asención-Delaney, 2014). They allow for an examination of co-occurring features in a particular corpus of texts in order to ascertain characteristics of a language or (sub)genre of said language in ways that would not be otherwise possible via a priori syntactic and phrasal criteria. MDAs group linguistic features that tend to co-occur at proportional or inversely proportional rates on dimension scales, or indices based on large quantities of linguistic variables that cluster together. Provided the corpus is properly composed of texts of the same genre via a situational analysis, they can allow for highly reliable information about linguistic phenomena in specific genres (Biber, 1985; 1988; 1992; 1993; Friginal, 2013).

To illustrate, based on frequency counts of linguistic features, face-to-face conversations have been found to contain high numbers of present tense verbs, contractions, and second-person

pronouns. Conversely, dense, official documents favor much higher numbers of nouns (Biber, 1988). Taken as a whole, the MDA can be seen as a powerful means of exploring hypotheses about the nature of a language and its genres.

2.1.6 Subsequent MDA Studies

MDAs have been used to examine a number of registers in world languages such as Korean (Kim & Biber, 1994), Somali (Biber & Hared, 1992), Portuguese (Berber Sardinha et al., 2014), as well as Spanish (Biber et al., 2006), substantiating the idea that these kinds of techniques first piloted on the use of English texts can then also hold relevance for other world languages. These studies revealed the same pronounced divisions in other languages as in English: spoken language favors present tense verbs and pronouns, whereas written language is denser with nouns (Biber, 1995). One area still ripe for discussion, however, is specific subgenres.

With regards to MDA studies that show higher levels of genre specificity, Cao & Xiao (2013) used multidimensional analysis to compare scientific abstracts written by L1 and L2 English users. Significant differences were found on several scales. That is, the L1 English users utilized more first-person pronouns, whereas the L2 English users employed fewer such pronouns and more abstract phrases to distance themselves from their claims. However, the abstracts used in the study in question were for hard science articles, rather than social science or humanities articles. Thus, more research on academic humanities registers elucidated by MDA is needed. Gardner, Nesi, & Biber (2019) used MDAs to discover the difficulties of texts and variation therein for university learners in different disciplines in English, in a way similar to

Hardy and Römer's (2013) work in the Michigan Corpus of Upper-Level Student Papers (MICUSP). MDA has also been used to explore the characteristics of new genres on the web (Titak & Roberson, 2013) and to ascertain features of proficient writing (Crossley et al., 2014a).

As for analyses of published academic writing, Biber and Finegan (1994) found significant variation in linguistic features used in Introduction, Methods, Results, and Discussion sections of biomedical research articles, offering further support for the idea that distinct discourse strategies may occur in each section of a research article. Gray (2015) compiled scholarly articles from reputable journals in history, applied linguistics, philosophy, chemistry, political science, and physics. She also grouped articles in each discipline as being quantitative, qualitative, and theoretical. Her subsequent MDA revealed strong levels of variation in the occurrences of linguistic features according to the discipline and research paradigm in which they occurred. One gap that remains is knowledge of the linguistic features prevalent in LCW generated via MDA.

Concerning MDAs in Spanish, one finds a handful of studies worthy of discussion. Parodi (2005), building on Biber's (1988) methods, conducted a study on a wide variety of genres present in Chilean university students' reading materials. The results suggested that requirements of distinct text types involve varying concentrations of linguistic features. Next, Parodi (2007) conducted an MDA and compared dimension scores throughout three registers represented in his corpus: oral interviews, textbook writing, and Latin American fiction. The results showed that the linguistic features of Latin American fiction more closely resembled those prevalent in oral interviews than in textbook writing. Venegas (2010), additionally, found that it was possible to simplify the MDA process to obtain commensurate results using fewer

linguistic features. The degree to which LCW more closely resembles textbook or fiction writing may pose one area worth investigating in both languages. Biber et al. (2006) found similar differences using a distinct tagging heuristic across a wide variety of Spanish-language text-types. Asención-Delaney and Collentine (2011) conducted an MDA on a small corpus of third and fourth-year university Spanish students' essays. They found that a wider range of verb conjugations was one indicator of proficiency. Asención-Delaney (2014) also performed an MDA on a corpus of specialist genres, including humanities writing, produced by graduate students in Spanish. However, to date, none of these studies has examined LCW exclusively on a large scale. Using an MDA on a collection of LCAs would provide a crucial level of detail in describing the register in ways that may not have been attempted previously. Specifically, situating LCW in relation to other academic genres as a means of explaining how and why it differs from other academic registers could provide an overarching set of considerations to keep in mind when developing LCW training materials.

One notable shortcoming of MDAs is concerned with the accuracy of the automated software in obtaining part of speech tags for the texts it automatically encounters. One study, for example, reported that the syntactic tagger used in many studies achieved only 67% accuracy when attempting to classify the word "like" as either a verb or a preposition, according to the context (Vine, 2011). One of the most commonly used taggers in MDA studies, the Biber Tagger, has also been shown to experience difficulties in classifying past participles, present participles, and tokens of "that" (Biber, 2000), requiring manual inspection and repair (e.g., Biber, 2000; Gray, 2010; Picoral, 2018). As such, additional approaches, such as the practice of

correcting automated syntactic diagrams, commonly referred to as “Fixtagging” (Biber, 2000), are warranted to remedy possible imperfections.

Overall, while norms in literary criticism have emerged from a variety of top-down and bottom up criteria, one particular issue is a lack of key understanding of phrases that repeat consistently and appear characteristic of the genre. In the following section, we review formulaic language, lexical bundles, and the potential they may hold in addressing the research questions of this dissertation.

2.2 Formulaic language antecedents of LBs

In the present section, we briefly review developments in the study of formulaic language and the ways in which they culminated in lexical bundles (hereafter “LBs”). They are operationalized as register-specific clusters of three or more words that often appear together in a corpus (Biber et al., 1999; Cortes, 2015). We will begin by discussing LBs’ antecedents before arriving at its present applications in English-language student (Cortes, 2002; 2004; Shin et al., 2018) and academic writing (Cortes 2004), humanities writing in English and Spanish (Cortes, 2008; Johnston, 2017; Lake & Cortes, 2020), as well as gaps to address in the present dissertation.

The history of LBs is largely informed by research on formulaic language from the past century. Cortes (2015) credits Jespersen (1933) with an early establishment of the dichotomy between “spoken formulae,” or phrases remembered as blocks, and “free expressions,” which are open to the infinite creativity of a human’s linguistic capabilities. Palmer (1933), a contemporary, may have coined the term “collocation,” within the context of teaching frequent

expressions. He used the term to refer to a multiword unit that must be learned as a whole rather than as isolated words. Firth (1957) then emphasized viewing words not by definition — a criterion by which it becomes easy to lose subtleties in words classified as synonyms of one another — but rather by the meaning implied by how they interact with surrounding words. In more recent history, knowledge of collocations was further solidified by Sinclair's (1991) early computer-assisted analysis of text corpora. Yet the source most closely credited with giving way to the notion of LBs is Altenberg's & Egg-Olofsson's (1990) corpus-driven work on newspaper writing using mainframe computers. The lists of highly repeated phrases generated by computer programs laid the groundwork for Altenberg's (1993) later work on repeated phrases in the London Lund Corpus, consisting of about 500,000 words. Despite these studies' empirically validated achievements, one issue later remedied by work on LBs (Biber et al., 1999) is that the lexical bundle is a more restricted, and potentially, more useful construct by learners by virtue of the frequency criteria discussed in the previous paragraph. LBs are thought to be among the most objective units of analysis currently available in the domain of corpus-driven research in that their formation depends entirely on the texts collected and the empirical method of identification used rather than the researcher's prior intuitions (Cortes, 2015).

Whereas N-grams refer to word sequences (Doddington, 2002), LBs differ from N-grams of three or more words in three key aspects: frequency, range, and genre specificity. That is, N-grams encountered in a particular corpus can consist of infrequent idioms (i.e., *feel under the weather*, *get a handle on*, *shooting fish in a barrel*) infrequently discussed places (*the Principality of Liechtenstein*, *the Norwegian Svalbard Archipelago*), or unique lexical expressions (*access to library resources*), etc. However, to count as LBs, these N-grams would

need to appear approximately 10-20 times per million words and in at least 5 different texts (Cortes, 2015). Realistically, said N-grams would need to be generated from genre-specific texts, i.e., idiom research papers in the case of the former and articles on relatively seldom-discussed, Germanic-language speaking European territories in the case of the latter, to boost their chances of becoming LBs.

2.2.1 Subsequent LB Studies in English

A number of studies on academic writing among university students suggest that students ought to learn the LBs specific to the textual genres with which they will engage in their major fields after graduating (Cortes, 2002; 2004; 2006; Chen & Baker, 2010; Shin et al., 2018; Shin & Kim, 2017). To this end, Cortes (2006) calls for more:

“analyses of the use of recurrent word combinations, such as LBs, by professional and novice writers in different academic settings [which] may provide evidence for better understanding of the language practices of certain academic communities and could help discover new ways of helping students become members of those communities” (p. 402)

Indeed, concerning genre specific LBs, descriptive studies of LBs in published academic writing have yielded insights in chemistry (Jalali, 2014), education (Candarli & Jones, 2018), telecommunications (Pan et al., 2016), civil engineering (Gilmore & Millar, 2018), as well as specific subsections of research articles in a variety of disciplines (Cortes, 2013). As for disciplinary specific research on LBs in LCW, Johnston (2017) managed to find examples of LBs in this field of study. Yet these LBs were found in analyses produced in English by L2

learners of English. As such, the findings may not be maximally compatible for insight on expert-level LCW.

2.2.2 Studies on LBs in published Spanish writing

Currently, one finds fewer studies on LBs in Spanish, but the field is growing. Butler (1998), for instance, investigated collocational frameworks in general language use. Applying more rigorous lexical bundle criteria, Tracy-Ventura, Cortes, and Biber (2007) compared general academic articles in the humanities to those occurring in sociolinguistic interviews using a functional and structural taxonomy. The results revealed that Spanish has vast inventories of LBs, only a few of which are common to spoken and academic genres. Importantly, the majority of bundles detected in the study were found to occur exclusively in either spoken or academic written Spanish. Cortes & Hardy (2013) examined the thematic context, or “semantic preference” and semantic prosody or the negative/positive/neutral context in which lexical bundles were used in English and Spanish academic writing. Pérez-Llantada (2014) also found that there is a large inventory of lexical bundles to be found in several academic registers in Spanish. As for genre-specific studies, Cortes (2008) found in parallel comparable corpora (Johansson, 2007) of English and Spanish academic history articles substantial evidence of written Spanish’s tendencies to manifest higher volumes of 4-word LBs than English. Lake & Cortes (in press) also examined the communicative functions of LBs in methods sections in English and Spanish.

Thus, as the field has seen a limited number of studies on LBs in LCW in English and Spanish, there is still ample work to be done on their role in LCW. While semantic prosody and

preference can reveal a great deal about a series of LBs (cf. Cortes & Hardy, 2013), complementary disciplinary specificity is warranted. While Lake & Cortes (2020) examined equivalent noun-phrase LBs in English and Spanish that occurred in the sentence subject position in literary criticism, history, and psychology writing, the specific functions performed by each LB remain to be examined. Warren (2006), for instance, found through think-aloud sessions that, when conceiving research projects, literary critics attempt to situate their topic within a paradigm. However, little is known about whether these critics draw from a recognized body of established disciplinary phrases. LBs could also help us understand more about what building blocks compose a successful, expert-level literary argument at the sentence level, as well as the functions they perform (cf. Cortes, 2013), or *communicative purposes*, which are a construct that we discuss in the following section.

2.3 Communicative Purposes

Earlier manifestations of genre theory conceived of “communicative purpose” as referring to the overall communicative goal of a document (Bhatia, 1997; Johns, 1997). In light of this dialogue, however, Askehave & Swales (2001) added that a single document may have several potential communicative purposes. A shopping list, for instance, may serve as a memory aid for trips to the supermarket for some, whereas for others, it may serve as a document that restricts its reader from buying anything extraneous. The scholarly conversation progressed, as a consequence, toward the assertion that a text belonging to one genre may contain a smaller segment belonging to another genre — a phenomenon known as an “embedded genre” (Bhatia, 2005; Biber & Conrad, 2009). One salient example of embedded genres, each with distinct

communicative purposes, is the prototypical scientific research article, which often contains Introduction, Methods, Results, and Discussion sections (a schematic known collectively as “IMRD”). Taking into account the Swalesian construct of the rhetorical move as a segment of discourse with a particular communicative purpose (Kanoksilapatham, 2007), such as to summarize prior research, indicate a gap therein, or to report on the intention of the researcher in a given study (Swales, 1990), it has been proposed that each section of the empirical, quantitative research article is composed of a wide reach of communicative purposes within each subsection (Kanoksilapatham, 2007; Cortes, 2013).

Attempts to identify communicative purposes in LCW have occurred in both English and Spanish. Tankó (2017) argued that LCAs lacked explicit cohesion markers between the statement of the article’s purposes and its outcomes. Further, in reference to the long sentences found in said abstracts, “the level of elaboration may even affect the communicative potential of the part-genre” (p. 53). Similarly, González Arías (2011) found that, in Spanish, 46% of the biology abstracts in his sample made the purpose of the study explicit, whereas only 20% of LCAs and 6% of history articles did so. As previously mentioned, to establish a taxonomy of strategies used by literary critics, one possible solution may lie in understanding LCW in terms of its intra-disciplinary strengths, rather than by diagnosing ways in which it deviates from the norms established by hard and social sciences.

Thus, although a noticeable amount of research on LCW exists, some questions still remain. Notably, while research on formulaic language sequences and their connections to specific communicative purposes has enriched understandings of other genres (Cortes, 2013; Chen, 2019), LCW, to the author’s knowledge, does not currently benefit from a comprehensive

study through this lens. As a possible solution to further delineating these ideas, let us discuss Qualitative Content Analysis as a means by which to fill this gap.

2.3.1 Qualitative Content Analysis

Qualitative Content Analysis (hereafter “QCA”) may refer to a number of analytical paradigms toward the collection of data in an attempt to arrive at deductions via approaches not necessarily tied to the outcome of positivist, statistical inferences. Rather, QCA emphasizes emergent patterns in a collection of texts. In the present study, QCA is operationalized as a mixed-methods approach proposed by Mayring (2014; 2015) to lay a foundation for a systematic categorization of data with criteria in place for interrater reliability. Mixed-methods approaches have shown empirical value for exploring conventions in general academic writing (Moreno & Swales, 2018; Bitchener, 2010; Bruce, 2008) as well as discipline-specific manifestations thereof (Atai & Falah, 2005; Stoller, Robinson & Jones, 2008; Kanoksilapatham, 2005; 2015).

To establish degrees of consistency and interrater reliability, QCA takes the following measures. After deciding on the unit of textual analysis and the context in which it is produced, 50% of the texts are analyzed qualitatively. During this initial phase, categories for textual phenomena are formulated via brief summarization of passages. After reaching the 50% mark, no new categories are formed. A master list of categories is then inductively combined into a more manageable set. Finally, independent raters are trained on using the criteria and are given some sample data to code. In this way, one may see QCA as similar to past qualitative research approaches, but with an emphasis on its rigor toward the establishment of interrater reliability coefficients.

Currently, there are two major strands of research using QCA. The first deals with such areas of medical research as tendencies in behaviors by patients with melanoma (Engeli et al., 2016), children with birth defects (Heider & Steger, 2014), benzodiazepine dependence (Liebrenz et al., 2015), and nicotine withdrawal (Liebrenz et al., 2016). Other scholars have used QCA to study the development of academic literacy and writing at a variety of grade levels. Concerning its use toward textbooks, Maier & Budke (2016), for instance, used QCA on German and British geography textbooks. They found that the textbooks used in both school systems do not encourage planning tasks for changes in geographical environments among students. Other studies deal with the research on German accounting textbooks (Hoffjan & Wömpener, 2006), development of E-learning in Germany (Fischer et al., 2014), multilingual publishing agendas among L1 German scholars (Schluer, 2014), codeswitching behaviors among German-English bilingual preschoolers in Cape Town (Terveen, 2013), flipped tutoring environments (Gunkel et al., 2018), extensive reading practice (Kreft & Viebrock, 2017), the teaching of strategic management in accounting textbooks (Karlen & Compagnoni, 2017), cultural competence in an L2 English classroom (Göbel & Helmke, 2010), and the teaching of academic writing in English (Voigt & Girgensohn, 2015; Brinkschulte et al., 2015; Girgensohn, 2016; Li & Ngai, 2018; Karlen & Compagnoni, 2017). However, none of these studies has examined published LCW.

One notable study that serves as an example of an uplink between QCA and underexamined, written genres and formulaic language comes in the form of Chen's (2019) analysis of the communicative purposes of lexical bundles (LBs) in a corpus of textbooks. In this study, the researcher generated a list of LBs from a million-word corpus of textbook chapters. These bundles were then classified according to their structure and function. Finally, the

researcher chose 10% of the textbook chapters in the corpus at random. These chapters were then analyzed for the dominant communicative purpose in passages whose boundaries were segmented via a text tiler that segmented the discourse into vocabulary-based discourse units. Using QCA, intercoder reliability coefficients were established in the identification of the communicative purpose of each vocabulary unit. Finally, the LBs present in each passage were measured to determine degrees of reliable correspondence between LBs and communicative purposes in textbook writing.

Taken as a whole, one sees an incipient degree of applied linguistics groundwork established for the study of LCW. It is hypothesized that a genre as nonformulaic as LCW could not be studied with such formulaic measures. Accordingly, further bottom-up examinations are needed to devise a suitable heuristic by which to research LCW genre conventions from a variety of perspectives

2.4 Summary: The motivation for the dissertation study

In summary, via an exploratory examination of communicative purposes in English- and Spanish-language literary criticism writing, this project will make an attempt to reveal unexamined components of the register, responding to calls in a number of prior studies for more transparency in how writing is achieved (Warren, 2006; Thiéme, 2017). As a consequence, new pedagogical materials could be based on these findings to enhance transparency in the teaching of the genre (e.g., Wilder, 2012). Since mixed-methods academic purposes linguistics studies have resulted in increased ability to teach academic writing generally for more scientific

disciplines (Swales & Feak, 2004; Swales, 1990; Robinson et al., 2008), it appears that similar boons to teaching humanities writing at the upper- and post-graduate levels await.

Given the state of LCW and its need for fresh, analytical approaches to the design of instructional materials, it appears that MDA, lexical bundle analyses, and QCA are worth performing on a corpus of texts belonging to this genre to determine whether any of the findings could enrich the quality of pedagogical materials for the field. These approaches, taken together, may lay the groundwork for a model by which to analyze other types of academic writing containing features distinct from those found in quantitative hard and social sciences. Accordingly, what will follow in the next section is a methodological proposal for a study that addresses the gaps outlined in the prior sections of this chapter.

This chapter's examination of prior studies shows that LCW has not currently had an exhaustive empirical study via the kinds of measures that have improved our understanding of other disciplines. As such, this dissertation examines first the kinds of linguistic features that are characteristically frequent and infrequent in the register via a multidimensional analysis (Research Question 1).

Moreover, given the enhanced teachability found as a consequence of lexical bundles that has resulted in insights in other disciplines (Cortes, 2008; Cortes, 2013; Pérez-Llantada, 2014; Lake & Cortes, in press), this dissertation takes recommended *a priori* situational criteria by which to arrive at a list of lexical bundles which may enhance the probability that learners can produce target-like sentences in LCW (Research Question 2).

Next, in line with previous findings on the communicative power of lexical bundles (Cortes, 2013; Chen, 2019; Lake & Cortes, in press), the dissertation entertains the possibility

that some lexical bundles may lend themselves uniquely to the realization of certain communicative purposes (Research Question 3). Finally, the dissertation also exemplifies the range of communicative purposes that can be identified in English and Spanish, as well as the possibility of variation in the two languages (Research Question 4).

To conclude, the examination of past work carried out in this chapter reveals that one finds a great deal of territory unexplored in LCW. Opening up new research avenues in these gaps could enhance the transparency of LCW and its disciplinary norms. In the event that new, identifiable features can be revealed through the approaches of MDAs, LBs, and communicative purposes, data that can be channeled into enhanced instructional materials for humanities majors may become available. This combination of theoretical and empirical orientations may, in turn, set forward the momentum necessary by linking applied linguistics and literary criticism to benefit novice literary critics who, much like novices from around the world in other disciplines (cf. Belcher, 2007), must publish their research in reputable journals.

3 CORPORA AND METHODOLOGY

The present chapter introduces the methods by which the three-pronged approach to descriptive analysis of literary criticism writing (LCW) was completed. Section 3.1 describes how the corpora of literary criticism were compiled with a number of focal criteria and, the register analysis for academic journal articles (Gray, 2015) and the variables it inspects. Section 3.2 explains the process by which multidimensional analysis (hereafter MDA) was carried out in English and Spanish to answer Research Question 1. Addressing Research Question 2, Section 3.3 identifies the structural and functional analysis processes performed on the lexical bundle dataset. Section 3.4 explains how communicative purposes were identified in the dataset of the present study, answering Research Question #3, in addition to the process of rater training and their interaction with the experimental paradigm to address research question 4. The chapter concludes with a summary of the methodological processes.

3.1 Corpora

The present study analyzed two comparable parallel corpora (Johansson, 2007) in Spanish and English, which met the following situational alignment criteria (Biber & Conrad, 2009). First, all articles came from literary criticism journals recognized as reputable and peer-reviewed by either the MLA International Bibliography, the Latin American journal validation database LATINDEX, or both. This step was important in light of rising for-profit academic journals, which have been found in the discipline of nursing, for example, to be of low caliber when compared to other journals that are more respected in a given field (Oermann et al., 2018). Next, text files in the corpora consisted of pieces published in literary criticism considered by the journals to be either articles or essays, rather than book reviews, notes, forum pieces, or other

publicity texts found in such journals. This methodological decision was taken since past research has found that the linguistic features in qualitative and quantitative studies in applied linguistics and political science, for example, differ significantly (Gray, 2015). Hypothesizing that such differences may exist in kinds of literary criticism journals as well, to ensure maximum comparability of texts in corpora, only articles examining phenomena in a single novel are included, meaning that comparative analyses and research on films, poems, short stories, auto/biographies, and plays were excluded. Two one-million-word corpora were compiled in line with the above criteria, as a million words is widely regarded as an acceptable quantity of text from which to be able to make generalizable claims about a variety of language (Biber & Conrad, 2001). Tables 3.1 and 3.2 show the country, journal, number of articles, and words represented from each journal by language.

Table 3.1. English corpus composition by country, journal, articles, and word count

Country	Journal	Initial words	Quotes removed	Count
Australia	Australian Literary Studies	73,601	57,359	10
Canada	Canadian Literature	162,944	132,324	24
Canada	Canadian Review of American Studies	154,249	121,923	19
South Africa	Literator	165,478	129,451	24
UK	Journal of Postcolonial Writing	156,310	127,805	24
USA	American Literature	175,794	139,590	16
USA	MELUS	250,308	202,481	25
USA	PMLA	94,709	75,409	10
	<u>Total</u>	<u>1,233,393</u>	<u>986,342</u>	<u>152</u>

Table 3.2. Spanish corpus composition by host country, journal, articles, and word count

Country	Journal	Initial words	Quotes removed	Articles
Argentina	Anclajes	97,096	77,442	13
Argentina	Orbis Tertius	89,473	72,191	14
Chile	Revista Chilena de literatura	157,004	124,464	19
Colombia	Literatura: Teoría, Historia, Crítica	89,098	68,935	11
Costa Rica	ISTMO	98,845	79,544	16
Mexico	Literatura Mexicana	148,785	110,745	20
Spain	Anales de Literatura Hispanoamericana	105,118	74,877	15
Spain	Revista de Filología Románica	113,657	83,228	20
Spain	Revista de literatura	105,562	81,138	10
UK	Bulletin of Hispanic Studies	149,740	111,237	20
US	Hispania	104,832	86,981	16
	<u>Total</u>	<u>1,259,210</u>	<u>970,782</u>	<u>174</u>

Columns 3 and 4 of Tables 3.1 and 3.2 represent word counts before and after direct quotes from primary sources were removed. This decision was taken in line with Cortes (2008), who removed direct quotes from history articles. The rationale behind this decision was to understand how historians write rather than the historical figures analyzed and quoted in the historians' articles. Similarly, this project aims to analyze how literary critics write, rather than how characters in novels talk when cited as evidence of a literary scholars' claims. However, only direct quotes containing verb phrases were deleted, leaving open the possibility for the inclusion of specialist terminology, which may often consist of noun phrases.

3.1.1 Register analysis

One heuristic used to identify disciplinary characteristics of journal writing is an analysis of the situational characteristics of texts of each article in a given corpus (Gray, 2015). To that end, Gray's specialized situational analysis framework for journal articles was used. More

general frameworks (i.e., Biber & Conrad, 2009) hold utility for comparing broadly different genres of texts to one another. However, since a great deal of variation in academic writing has been found from discipline to discipline (Gray, 2015), a more specialized framework assists specialists in understanding the finer nuances of journal articles in specific disciplines. Gray's (2015) framework allows academic writing researchers to consider the following variables (shown in Table 3.3 with brief explanations) for each article in their corpora:

Table 3.3. Situational characteristics framework (Gray, 2015)

Category	Variable	Explanation
1. Participants	A. Writer	The number of authors listed in an article.
	A. Length	The number of pages in a given article. Once data is available for a whole corpus, mean page length, standard deviation, and range are reported.
	B. Headings	Considers whether section headings (numbered or un-numbered) are used within the text.
	C. Use of Abstracts	Determines whether or not the article contains an abstract.
	D. Visual Elements	Considers whether an article uses any tables, figures, or equations.
2. Textual Layout & Organization	E. Sections / Organization	Considers to what extent the article follows a standardized section heading pattern of indicating an introduction, conclusion, Introduction-Methods-Results-Discussion, or some other schematic.
3. Setting	A. Nature of Journal	The degree of specialization in each journal from which articles have been selected.
4. Subject/Topic	A. General Topic of the Discipline	The superordinate topical category explored in a given discipline.
	A. General Academic Purpose	The general purpose of studies in a given discipline.
5. Purpose	A. Presence of Observed Data	Whether or not data appears in the research article.
	B. Use of Numerical Data	Whether or not said data is quantitative.
6. Nature of Data or Evidence		

Table 3.3 Continued

	C. Primary Presentation of Evidence	Asks whether data is primarily displayed via prose descriptions, quantitative tables, or mathematical formulae.
6. Nature of Data or Evidence	D. Object of Study Varied (based on trends in the corpus)	Items studied, generally speaking, in observed trends in the corpus.
	A. General Method Type	Asks whether data is observational or experimental in nature.
7. Methodology	B. Statistical Techniques	Whether or not any statistical tests are used in analysis and whether these are descriptive or more advanced.
	A. Explicitness of Purpose	Whether or not an article states its purpose in the abstract or the first few pages.
	B. Explicitness of Research Questions	Whether or not research questions are stated.
	C. Explicitness of Citations	The presence of citations as integrated or primarily relegated to footnotes.
	D. Explanation of Evidence	Whether or not evidence is described at length as being evidence, metadiscursively speaking.
8. Explicitness of Research Design	E. Explanation of Procedures	Whether or not the procedures by which evidence was gathered were explained in any great detail.

After the articles were collected, they were classified before interpreting the MDA in preparation for its results. 153 articles were analyzed in English and 175 in Spanish using the above framework.

3.2 Multidimensional Analysis (MDA) of linguistic features

To explore Research Question 1, which seeks to examine what an MDA of English and Spanish-literary criticism can tell us about how these texts appear in relation to other genres, an MDA of linguistic features occurring in the English and Spanish sub-corpora was performed.

What follows are explanations of the unique processes carried out to obtain dimensional scores for the text files in both languages. Dimension scores are placed in conversation with prior MDAs in English and Spanish to determine how Literary Criticism Writing compares to other genres.

3.2.1 MDA in English

As mentioned in the previous chapter, in descriptive analyses of text genres, Dimension Scores contain information generated from large numbers of linguistic features that tend to co-occur at predictable rates. While some multidimensional analyses make their own intra-corpus dimension scores (e.g., Asención-Delany & Collentine, 2011; Asención-Delaney, 2014; Gray, 2015; Berber Sardinha, 2019), others (e.g., Berber Sardinha & Veirano Pinto, 2017) place the findings of a given corpus in dialogue with those norms obtained by Biber (1988). In his original study, Biber counted linguistic features with software which he used on the London-Lund corpus, a collection of texts from a wide variety of genres, ranging from official documents to romance novels. The present study opts to dialogue with Biber's (1988) lexicogrammatical profiles for text genres in English, since, to the author's knowledge, no other study exists which puts literary criticism writing in conversation with other genres using a multidimensional analysis approach.

Studies that do compare their findings with Biber's may consider dimension scores based on approximately six dimensions, which are described in Table 3.4.

Table 3.4. Description of Biber's (1988) six dimensions and associated linguistic features

Dimension	Description
1: Involved vs. Informational Production	Texts with lots of first- and second-person pronouns and fewer nouns (i.e., conversations) are more involved, whereas those with more nouns and fewer pronouns are more informational in nature.
2: Narrative vs. Non-Narrative Concerns	Texts with more past tense verbs and third-person pronouns are more narrative. A text with fewer of these and other features is more non-narrative.
3: Explicit vs. Situation-Dependent Reference	Assesses the amount of contextual information present in a given text. Texts with more wh- relative clauses in subject and object position, as well as nominalizations, are more explicit. Texts with more adverbs, time adverbials, and place adverbials are more situation-dependent.
4: Overt Expression of Persuasion	Texts which use more prediction and necessity modals in conjunction with infinitives, as well as suasive verbs and split auxiliaries, are more overtly persuasive, whereas those with fewer such characteristics are less overtly so.
5: Abstract vs. Non-Abstract Information	More abstract texts have more agentless and By-passive voice tokens, as well as more conjuncts and past participial clauses, whereas those with fewer such features are closer to the Non-abstract end of the continuum.
6: On-line Informational Elaboration	Assesses the amount of demonstratives, <i>that</i> clauses as verb complements, <i>that</i> clauses as adjective complements, as well as <i>that</i> relative clauses in object positions.

It is beyond the scope of this project to describe how all the dimension scores are calculated. However, to allow a fair degree of transparency in this methods chapter, what follows now is an in-depth explanation of how dimension scores for *Dimension 2: Narrative Concerns vs. Non-Narrative Concerns* (Biber, 1988) are calculated using an abstract from one of the articles in our English-language corpus.

To begin, scores for *Dimension 2, Narrative vs. Non-Narrative Concerns* are computed based on the following linguistic features:

- Past tense verbs

- Third-person pronouns (excluding “it”)
- Perfect aspect
- Public verbs

Instances of the above features appear bolded in Figure 3.1.

“This article **suggests** that the proliferation of garbage in Philip K. Dick’s *Martian Time-Slip* is not descriptive or predictive but aesthetic. The use of trash in literary and visual art would have **been** familiar to Dick from the works of Beat poets and assemblage artists circulating in the San Francisco area during the late 1950s and early 1960s, but the specific garbage aesthetic of the novel is based on two principles: the accumulation of found objects and the sacred grotesque, including its suggestion of renewal by inversion. The first principle is inseparable from Dick’s analysis of consumer capitalism, while the second highlights the playful inversion of religious imagery and the possibility of renewal arising from the materiality, debasement, and incoherence of trash. While Dick anticipates postmodern trash aesthetics, especially in the framing of accumulated found objects, **his** emphasis on the sacred grotesque distinguishes **his** garbage texts from the “palimpsestic” representations of garbage **described** by Robert Stam and others.” (472.txt)

Figure 3.1. Abstract of The Aesthetics of Garbage in Philip K Dick’s Martian Time Slip

To calculate a Dimension 2 score for the abstract found in Figure 3.1, the Biber Tagger first uses some probabilistic models in concert with pre-established dictionary database files to add syntactic information about each word in the above passage. This process is called Part-of-Speech (hereafter “PoS”) Tagging. To illustrate, the final clause found in Figure 3.1, PoS tagged by the Biber Tagger, appears in Table 3.5.

Table 3.5. Example of a tagged sentence

Tokenized form	PoS Tag	[PoS Tag Meaning]
his	^pp\$+pp3+++	[Possessive pronoun, third-person]
emphasis	^nn++++	[singular noun phrase]
on	^in++++	[preposition]
the	^ati++++	[Determiner]
sacred	^jj+atrb++xvbnx+	[Adjective, attributive, past participle?]
grotesque	^jj++++	[Adjective]
distinguishes	^vbz++++	[third-person present tense verb]
his	^pp\$+pp3+++	[Possessive pronoun, third-person]
garbage	^nn++++	[singular noun phrase]
texts	^nns++++	[plural noun phrase]
from	^in++++	[preposition]
the	^ati++++	[determiner]
palimpsestic	^jj+atrb+++	[Adjective, attributive]
representations	^nns+nom+++	[Plural noun, suffix-derived noun]
of	^in++++	[preposition]
garbage	^nn++++	[singular noun phrase]
described	^vwnb++by+xvbn+	[passive postnominal modifier, by-passive]
by	^in++++	[preposition]
Robert	^np++++	[Proper Noun]
Stam	^np+++??+	[Proper Noun, ambiguous]
and	^cc++++	[coordinating conjunction]
others	^aps++++	[“others”]
.	^zz++++	[Punctuation]

The first column is the tokenized form of a word, meaning that the word has been stripped of punctuation and other extraneous features, rendering it ready for comparison with the Biber Tagger’s internal dictionary files. The PoS tag is the Biber Tagger’s rendering of available syntactic information about a given word. Of interest to us are the tokens of “his,” which have the “pp3” PoS tag, standing for “personal pronoun 3,” indicating that they will be included in counts of “third-person pronouns.” The “vwnb” tag applied to the instance of “described” signals

that it is a kind of past tense verb. As such, the token forms part of the of past tense verb total. The verb phrase “have been” is added to totals of “perfect aspect.” Finally, “suggests” registers as a public verb, meaning that it refers to a publicly observable process, usually with an audience, according to the Biber Tagger database files.

After all files in the corpus are PoS tagged, a second program called TagCount counts all linguistic features based on predefined patterns and outputs a spreadsheet with around 170 columns, 51 of which are used in comparative MDAs looking at Dimensions 1-6 of Biber’s (1988) analysis). Table 3.6 lists the raw frequency counts for the four Dimension 2 linguistic features which appeared in bold in Figure 3.1.

Table 3.6. Step 1: Obtaining frequency counts from which to calculate dimension scores

Feature	Word	Tokens
Past tense verbs	described	1
Third-person pronouns	his	2
Perfect aspect	been	1
Public verbs	suggests	1

The next step in dimensional score calculation is to make frequency counts obtained from a document comparable to counts in other files. Counts in studies based on Biber’s (1988) work normalize frequency counts to occurrence rates per thousand words. This step is deemed necessary to maximize the potential for comparability between text files of varying lengths. Think of the classic math problem type that tells you the rate at which a train traversed from Atlanta and St. Louis to estimate its travel time to San Francisco. The assumption is similar. Keeping in mind that the abstract in Figure 3.1 consists of 158 words, if it uses past tense verbs

once in that amount of text, we use the following formula to obtain an estimate of how many past tense verbs would occur in a unit of 1,000 words:

$$\frac{\text{Raw frequency count of linguistic feature} \times 1000}{\text{Word count of the document chosen for analysis}}$$

With the above formula, multiplying the counts in for each feature by 1,000 and then dividing that number by 158, the word count of the abstract shown in Figure 3.1, we obtain the following normalized frequency counts (i.e., assuming all things equal, the features would have occurred at the same rate in a document of 1,000 words) for linguistic features in the above abstract. Table 3.7 contains said normalized frequency counts.

Table 3.7. Step 2: Calculating frequency counts normalized to occurrences per thousand words

Feature	Word	Frequency	((Freq. × 1000) / 158)
Past tense verbs	described	1	6.3
Third-person pronouns	his	2	12.7
Perfect aspect	been	1	6.3
Public verbs	suggests	1	6.3

Now the frequency counts are ready to be put in dialogue with Biber's findings to see how they score within the parameters of established lexicogrammatical profiles for English-language text genres.

All Dimension Scores computed within the parameters of Biber (1988) are calculated in relation to standardized scores based on the means and standard deviations obtained from the million words of text from various genres in the Lancaster-Oslo-Bergen Corpus of British English (Johansson et al., 1978, Johansson, 1982) and the 500,000-word London-Lund Corpus of Spoken English (Svartvik & Quirk, 1980; Johansson, 1982). Table 3.8 displays our normalized scores generated, along with the means and SDs that Biber (1988) reported for Dimension 2 linguistic features.

Table 3.8. Step 3: Comparing normalized frequencies with means and SDs for linguistic features found in Biber (1988)

Feature	Our Abstract	Corpus-level Mean (Biber, 1988)	Corpus-level SD (Biber, 1988)
Past tense verbs	6.3	40.1	30.4
Third-person pronouns	12.7	29.9	22.5
Perfect aspect	6.3	8.6	5.2
Public verbs	6.3	7.7	5.4

The next order of business is to obtain z-scores for the frequency counts of each linguistic feature analyzed in a given dimension. Z-scores are standardized scores which can communicate quickly to analysts whether a number is typical or atypical according to pre-established means. Specifically, they state how many standard deviations above or below a predefined mean a number is. In this case, our z-scores are calculated by comparing the normalized frequency counts of a linguistic feature in relation to the mean and SD of said feature obtained by Biber (1988) using the following formula:

$$\frac{\text{Raw frequency count of a linguistic feature} - \text{Mean score for linguistic feature in Biber (1988)}}{\text{Standard deviation of linguistic feature in Biber (1988)}}$$

To prepare counts for processing with the above formula, all variables are rounded to the nearest whole number. Table 3.9 shows Z-scores for our linguistic features calculated after rounding is applied to all numbers.

Table 3.9. Summary of Dimension 2 calculation procedures for the abstract in Figure 3.1

Feature	Fig. 3.1	Biber Mean	Biber SD	Z-Score
Past tense verbs	6	40	30	-1.133
Third-person pronouns	13	30	23	-0.739
Perfect aspect	6	9	5	-0.600
Public verbs	6	8	5	-0.400
			<u>Dimension 2</u>	<u>Sum: -2.872</u>

Thus, applying the above formula to our counts for past tense verbs, we obtain the following value: $((6 - 40) / 30) = -1.133$. Note, for example, that the amount of past tense verbs in Figure 3.1 is more than 1 standard deviation below the mean frequency of past tense verbs, whereas the number of public verbs in Figure 3.1 is within less than 1 standard deviation of the mean for public verbs. Summing all of the z-scores of all linguistic features calculated in the same fashion by subtracting the mean from the feature count and dividing by the standard deviation, we obtain a Dimension 2 score of -2.872. To understand the significance of this Dimension 2 score, Figure 3.2 shows how the score compares to Dimension 2 Scores obtained for other documents (Biber, 1988).

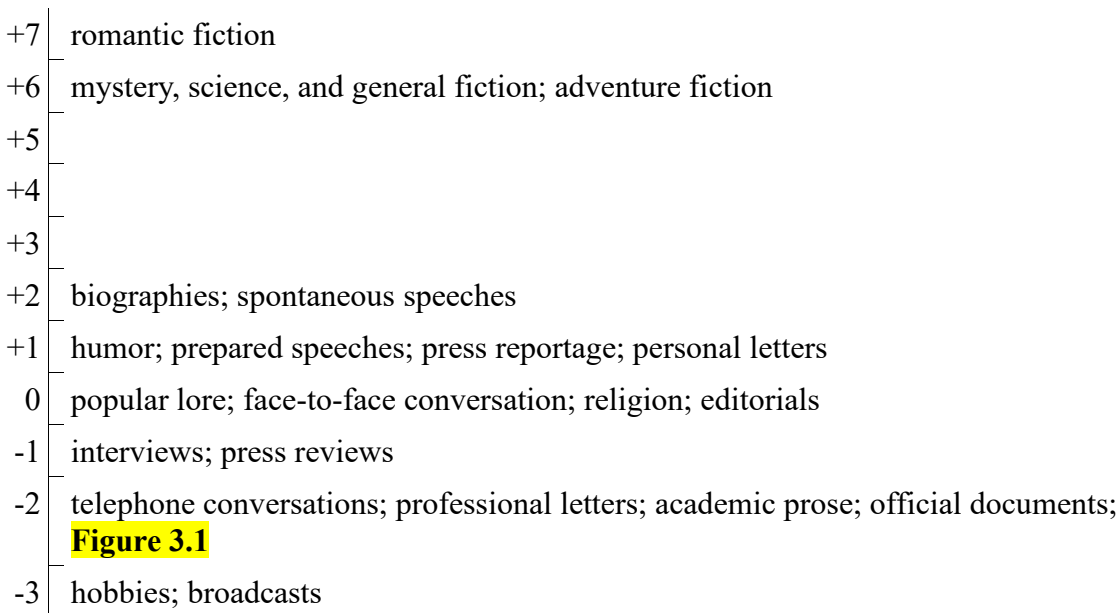


Figure 3.2. Mean Scores for Figure 3.1 and each genre in Biber (1988) on Dimension 2: Narrative Versus Non-Narrative Concerns

As shown in Figure 3.2, fiction in several different genres uses past tense verbs and related features at rates 6 standard deviations above the mean for the rest of the English-language

documents analyzed in the London-Lund corpus (Biber, 1988). Academic prose, however, uses Dimension 2 features at about 2-3 standard deviations below the mean, which is where our abstract in Figure 3.1 falls. This information tells us that the abstract is indeed on par with norms reported for other passages of academic prose, insofar as narrating events in the past is concerned. The abstract in Figure 3.1, moreover, appears less narrative than all types of fiction, biographies, and interview transcripts. While not devoid of past tense features, it simply uses them noticeably less than the mean levels in the London-Lund corpus (Biber, 1988). These findings suggest that the text in Figure 3.1, using very few instances of the past tense, resembles genre characteristics found to be common in other passages from academic prose and official documents. The above process uses the same principles for the calculation of Dimensions 1, 3, 4, 5, and 6. More remarks on salient linguistic features in these dimensions are available in Chapter 4: Multidimensional Analysis Results.

3.2.1.1 Fixtagging

The Fix-Tagger, a tool well-attested in MDA studies (García & Asención, 2001; Csomay, 2006; King, 2007; Reppen, 2007; Biber & Gray, 2013), allows linguists to validate the syntactic metadata for text files that is automatically produced by the Biber Tagger (Biber, 1988).

Fixtagging is an important process for ensuring the quality of the data that goes into an MDA. One study reported that the Biber Tagger achieved only 67% accuracy when attempting to classify the word “like” as either a verb or a preposition, according to the context (Vine, 2011). Currently, there are three versions of the Fix-Tagger developed in the computer programming languages Delphi (Biber, 2000), Python (Gray, 2010), and JavaScript (Picoral, 2018). The

present study used a modified version of the Python version, in which hexadecimal code was edited by the researcher to make the text appear at a more accessible font to reduce eyestrain.

In the present study, not all tags were inspected. Only those in syntactic categories that appear the most challenging to disambiguate for the Biber Tagger were inspected (Biber, 2000): past participles, tokens of “that,” and present participles. Predefined regular-expression (or “regex”) filters in the software allow it to search through only these features. Post-hoc corrections are then made for nominalizations, which have been found to need additional assistance in detection due to disciplinary specificity of technical terminology (Gray, personal communication, 2019).

Table 3.10 shows the word number from the above abstract discussed “The Aesthetics of Garbage...”. The left column represents the word in chronological order from the abstract. The next column shows a human-readable word. The next column, “Old Tag,” shows the tag initially assigned to each word. When an item to fixtag was found, the tag was changed. Any such changes are indicated in the “Fixed Tag” column.

Table 3.10. Comparison of the abstract tagged automatically by the Biber Tagger and tags fixed manually by the researcher.

# Word	Old Tag	Fixed Tag
97 the	ati++++	
98 accumulation	nn+nom+++	
99 of	in++++	
100 found	vbd+vprv++xvbn+	jj+atrb++xvbnx+
101 objects	nns++++	
[...]		
103 the	ati++++	
104 sacred	jj++++	jj+atrb++xvbnx+
105 grotesque	jj++++	
[...]		
107 including	in++++	vbg+nf++xvbg+
108 its	pp\$+it++++	
109 suggestion	nn+nom+++	
110 of	in++++	
111 renewal	nn++++	nn+nom+++
[...]		
155 While	cs+sub+++	
156 Dick	np++++	
157 anticipates	nns+nom+++	vbz++++
158 postmodern	jj+atrb++xvbn+	
159 trash	nn++++	
160 aesthetics	nn++++	
161 ,	zz++++	
162 especially	rb++++	
163 in	in++++	
164 the	ati++++	
165 framing	jj+++xvbg+	nvbg+++xvbgx+
166 of	in++++	
167 accumulated	nn++++	jj+atrb++xvbnx+
168 found	vbd+vprv++xvbn+	jj+atrb++xvbnx+
169 objects	nns++++	
[...]		
184 representations	nns+nom+++	

Table 3.10 Continued

185 of	in++++	
186 garbage	nn+nom+++	nn++++
187 described	vwbn+++xvbn+	vwbn++by+xvbn+
188 by	in++++	
189 Robert	np++++	
190 Stam	np+++??+	

Observe, for instance, word #100. In the noun phrase “found objects,” the instance of “found” was initially interpreted by the Biber Tagger as a past-tense verb, rather than an attributive adjective. Accordingly, the tag “vbd+vprv++xvbn+” was changed manually to “jj,” meaning “adjective,” and “atrb,” meaning “attributive. The “atrb” component was also added to “sacred,” due to the Fixtagger’s filter marking words ending in -ed as past tense verbs.

Next, while “including” was initially treated as a preposition with the “in” tag, the researcher chose to change it to “vbg+nf,” indicating that the present participle token initiates a nonfinite present participial clause. Note also that “renewal” was detected in post-hoc inspections to qualify as a nominalization.

One also sees in Word #157 that “anticipates” was initially interpreted as a nominalized plural noun due to the predefined rulesets for identifying less-frequent words in the Biber tagger. However, the post-hoc nominalization inspection process also removes false positives (Gray, personal communication, 2019). #165, “framing,” first treated as an adjective, was corrected to contain the tag “nvbg,” indicating that it is a noun phrase consisting of a present participle. Note also the noun phrase in words 167-169, “accumulated found objects,” with the first two words tagged beforehand as a noun and a past tense verb. They were retagged as attributive adjectives modifying the noun “object.” Also, observe the phrase “representations of garbage described by

Robert Stam.” “Garbage” was treated as a nominalization, due to its ending in the letters “age.” However, nouns using these suffixes with unclear relationships to verbs that have been lost in etymological history are not often regarded as nominalizations in fixtagged MDA studies (Gray, personal communication, 2019). Additionally, the past tense verb “described,” as it is followed by the preposition “by” with an explicit agent, was repaired to include the tag “by.” Finally, note that “grotesque” in row #105 was tagged as an adjective with the “jj” tag when it ought to be a noun. That such cases elude the current fixtagging practices is acknowledged as an imperfection of the research methodology.

Although the researcher also modified the JavaScript version of the English-language Fixtagger (Picoral, 2018) to work with the Spanish tagset and painstakingly fixed the tags with present participles, past participles, and tokens of *que* (a common, grammatical Spanish-language counterpart of “that”) as well as nominalizations, a highly cited Spanish MDA study (Biber et al., 2006) does not report means or standard deviations for its linguistic features, rendering impossible the ability to dialogue quantitatively with it to the same degree as one can with Biber (1988). The following section discusses a secondary measure taken in the form of an MDA study with fewer linguistic features in Spanish used to assess the properties of writing in several genres (Venegas, 2010).

3.2.2 MDA in Spanish

In lieu of the ability to Dialogue with another Spanish MDA Study (Biber et al., 2006), The present study opted to use Venegas’ (2010) methodological parameters for Dimensional Density Scores obtained from Spanish texts, which are calculated via analytical techniques for

measuring variation in linguistic features by genre. To begin, using supplemental documents graciously provided by Venegas (personal communication, 2020), a SiNLP dictionary file (Crossley et al., 2014b) was compiled to obtain counts for the following linguistic features.

SiNLP is a software program that, when given a user-specified dictionary file of features to count in a folder of text files, creates a spreadsheet with the number of said features in each text file. It has been used to measure cohesive features in high school students' writing (Crossley et al., 2014b) and lexical bundles in medical writing (Mbodj & Crossley, 2020), among other features.

The SiNLP dictionary file contained criteria for measuring the following features in Spanish:

- Cause/effect adverbial clauses
- Time adverbs
- Negation adverbs
- Second-person singular subject pronouns
- First-person singular subject pronouns
- Second-person plural subject pronouns
- Plural pronouns
- Private verbs
- First-person pronouns
- Adverbs of manner
- Modal verbs of obligation
- Nominalizations

Frequency counts for each of those features were obtained using SiNLP. The remaining frequency counts based on lemma-based and pattern-based syntactic criteria were obtained using the Spanish Tagger and Spanish Tag Count (Biber, 2006) software tools in conjunction with regular expression queries in Geany, an open-source text editing program that allows for tabbed browsing of hundreds of text files at once (Tröger et al., 2019). The features for which counts were obtained using these three programs were:

- Periphrastic future
- Imperfect past

- Preterit
- Modal verbs of volition
- First-person singular inflections
- Active forms of the verb *ser* (to be, durative)
- Hedging verbs
- Modal verbs of possibility
- Predicative adjectives
- Subjunctive mood
- Participles in adjectival functions
- Prepositional phrases as noun complements

Table 3.11 shows examples of each linguistic feature included in an analysis of Dimension Density scores in Spanish, along with the factor loadings reported by Venegas (2010).

Table 3.11. Spanish MDA features with examples, loadings used by Venegas (2010)

Dimension	Feature	Examples	Loadings
1: Contextual and Interactive Focus	Cause / effect adverbial clauses	porque, así, debido, ende, ya que, puesto que, por lo tanto, de este modo [because, thus, due, consequently, since, given that, therefore, in this way]	0.945
	Time adverbs	pronto, temprano, siempre, nunca, jamás, todavía, aún, ya, ahora, luego, tarde, ayer, hoy, anteayer, mañana [soon, early, always, never, nevermore, still, yet, already, now, later, late, yesterday, today, the day before yesterday, tomorrow]	0.934
	Negation adverbs	no [no]	0.928
	Second-person singular pronouns	tú [you]	0.911
	First-person singular pronouns	yo [I]	0.823
2: Narrative Focus	Second-person singular pronouns	tú [you]	0.842
	First-person singular pronouns	yo [I]	0.828
	Periphrastic future	ir + a + infinitive [to go + to + infinitive]	0.823

Table 3.11 Continued			
	Imperfect past	imperfect past tense e.g., escribía [I/he/she/it used to write]	0.820
	Plural pronouns	ellas, ellos, los, las, les, mis, sus, tus, Uds./ustedes [they (fem.), they (masc.), them (masc.) & them (fem.) (direct object), them (indirect object), my (plural), their/her/his/its (plural), your (plural), you (formal, plural)]	0.708
3: Commitment Focus	Private verbs	descubrir, creer, concluir, decidir, adivinar, encontrar, sentir, temer, determinar, demostrar, estimar, reconocer [to discover, to believe, to conclude, to decide, to guess, to find, to feel, to fear, to determine, to demonstrate, to estimate/regard, to recognize]	0.824
	First-person singular pronouns	me, mí, conmigo, migo, yo [me, to me, with me, I]	0.789
	Preterit	Simple past tense	0.705
	Modal verbs of volition	querer [to want] + infinitive	0.655
	First person singular inflections	escribo, escribí, escribiré, etc. [I write, I wrote, I will write]	0.640
	Active forms of the verb <i>ser</i> (to be, durative)	<i>ser</i> + noun; <i>ser</i> + adjective; <i>ser</i> + pronoun; <i>ser</i> + determiner	0.671
4: Modalizing Focus	Hedging verbs	<i>parecer que, creer que, estimar que,</i> <i>pensar que</i> [to seem that, to believe that, to estimate/regard that, to think that]	0.656
	Modal verbs of possibility	<i>poder</i> [to be able] + all parts of speech except for nouns	0.641
	Adverbs of manner	<i>-mente</i> [-ly]	0.606
	Predicative adjectives	<i>Estar / ser</i> [both aspects of the verb to be] + adverb; <i>estar / ser</i> [both aspects of the verb to be] + adjective	0.565
5: Informational Focus	Modal verbs of obligation	all conjugations of <i>deber</i> [should/ought/to owe]	0.496
	Subjunctive mood		0.494

Table 3.11 Continued

Nominalizations	<i>-ción, -ciones, -sión, -siones, -miento, -mientos</i> [-tion(s); -sion(s), -ment(s)]	0.456
Participles in adjectival functions	e.g., <i>experiencia vivida</i> [lived experience]	0.413
Prepositional phrases as noun complements	noun + preposition + noun; adjective + preposition + noun	0.413

Next, the following formula is used to calculate dimensional density scores.

$$ID_dT = \frac{\sum(fR_{1..n} \cdot \alpha_{1..n}) \cdot \sum(R)}{\sum pT}$$

A further explanation of how the formula plays out will be demonstrated using the abstract from the following Spanish-language literary criticism article:

- Corona, G. L. (2011). Infancia y opresión en *Balún Canán* de Rosario Castellanos. La niña como eje temático y estructural de la novela. *Bulletin of Hispanic Studies*, 88(7), 777-794.

Dimension 1 features are highlighted in blue, Dimension 4 in green, and Dimension 5 in yellow. No features for Dimensions 2 or 3 occurred in the abstract.

A visual inspection of the annotations in Figure 3.3 reveals that Dimension 5 features, highlighted in yellow, predominate. Venegas (2010) found that research articles possess a high amount of informational density relative to other text files.

Infancia y **opresión** en Balún Canán de Rosario Castellanos. La niña como eje temático y estructural de la novela

Resumen: Balún Canán (1957), de Rosario Castellanos, ha sido estudiada **fundamentalmente** con un enfoque feminista, y a veces también desde un **punto de vista** indigenista, pero falta hacer énfasis en el papel que desempeña la protagonista, la **niña sin nombre**, en la novela. En este artículo se tratará de abordar este aspecto, considerando a la niña como el eje temático y estructural de la **organización** novelística. En un primer nivel, el libro explora la **expulsión** de la niña del paraíso infantil. En un nivel más profundo, la niña representa la **opresión** social **no** sólo de los niños, sino también de las mujeres y los indios. Este análisis pondrá de manifiesto la lógica estructural que se halla detrás del uso de las diferentes voces narrativas en las partes I y III (primera persona) y en la parte II (tercera persona) de la novela. (19.txt)

[Childhood and **oppression** in Rosario Castellanos' Balún Canán. The girl as a thematic and structural axis of the novel

Abstract: Balún Canán (1957), by Rosario Castellanos, has been **primarily** studied from a feminist approach, and also sometimes from an indigenist **perspective**, but there is a dearth of emphasis made on the role played by the protagonist, the **nameless girl**, in the novel. In this article, an attempt will be made to address this aspect, considering the girl as the thematic and structural axis of the novel's **organization**. On a superficial level, the book explores the girl's **expulsion** from the paradise of childhood. On a deeper level, the girl represents the social **oppression** **not** only of children but also of women and indigenous peoples. This analysis will highlight the structural logic found behind the use of the different narrative voices in parts I and III (first-person) and in part II (third-person) of the novel.]

Figure 3.3. Spanish Abstract with dimensional features highlighted.

Table 3.12 summarizes the process by which the above detected linguistic features are counted and factored into dimensional density scores.

Table 3.12. Summary of Spanish DDS calculation process

Dimension	1	2	3	4	5
Features found in Figure 3.1	Negation adverb	N/A	N/A	Adverbs of manner	Nominalizations; prepositional phrases as noun complements
Tokens	no (1) [no]			<i>fundamentalmente</i> (1) [fundamentally]	<i>expulsión</i> [expulsion] (1), <i>opresión</i> [oppression] (2), <i>organización</i> [organization] (1); <i>niña sin nombre</i> [nameless girl] (1), <i>punto de vista</i> [point of view] (1)

Table 3.12 Continued

Feature tokens multiplied by Venegas' (2010) factorial weights	$1 * 0.928 = \underline{\mathbf{0.928}}$	$1 * 0.606 = \underline{\mathbf{0.606}}$	$4 * 0.456 = \underline{\mathbf{1.824}};$ $2 * 0.413 = \underline{\mathbf{0.826}};$
Product multiplied by the total number of features known for each factor (5)	$0.928 * 5 = \underline{\mathbf{4.64}}$	$0.606 * 5 = \underline{\mathbf{3.03}}$	$1.824 * 5 = \underline{\mathbf{9.12}};$ $0.826 * 5 = \underline{\mathbf{4.13}};$ $4.13 + 9.12 = \underline{\mathbf{13.25}}$
Sum divided by total words (161) in documents	0.029	0.019	0.082
DDS	0.029	0	0
		0.019	0.082

First, all words meeting lexicogrammatical criteria are added to total frequency counts for said features. Those frequency counts are then multiplied by established factorial loadings (shown in Table 3.11) for each feature (Venegas, 2010). The product of this calculation is then multiplied by the number of factors associated with a given dimension, in the case of this study, 5 per dimension. After each linguistic feature has gone through this process, the resulting products of any feature on the same dimension are added together. The resulting number is then divided by the total number of words in the document, in the case of this example, the 151 words for the title and abstract included in Figure 3.3. The resulting dimensional density scores for Figure 3.3 appear in Table 3.12.

Following the lead of the examples for the English MDA, we consider the Dimension 2 score placed in dialogue with Dimension 2 in Venegas (2010), which also assessed narrative features, appearing in Figure 3.4.

DDS Genres
0.250 Latin American Literature
0.200
0.150
0.100 Oral Didactic Corpus
0.050
0.000 PUCV-2006 Corpus; Public Politics; Scientific Research Articles; Figure 3.3

Figure 3.4. The abstract from Figure 3.3 classified according to its Dimension 2 features

With a Dimension 2 score of 0, it can be inferred that the abstract in Figure 3.3 is highly characteristic of scientific research articles, as well as political speech, due to its dearth of past tense verbs and other narrative features. The importance, however, of examining LCAs from additional linguistic angles will be discussed in Chapter 4, Multidimensional Analysis of English and Spanish.

3.3 Lexical Bundle generation

Lexical bundle lists were generated using AntConc (Anthony, 2014) with a frequency cut-off point of 20 times per million words occurring in at least 5 distinct texts, generally produced by distinct authors, (Biber et al., 2004) from both corpora. LBs meeting frequency and range criteria were collected from the corpora successfully. Tokens were individually inspected to remove cases of false positives and bundles interrupted by punctuation. Tokens of each bundle with intra-phrasal punctuation were excluded. For example, take the case of the LB *at the same time as*. The five words in the bundle do occur in the following extract from an article “**At the**

same time, as Marrow powerfully evidences...” However, accepted lexical bundle practice eliminates punctuation marks. For example, commas are classified as indicators of the insertion of new syntactic constructions containing different linguistic features. Thus, examples of “at the same time, as” were not considered tokens of the said bundle. Only clearer cut cases, such as the following were included: “ In this way the text unsettles the authority of readers to pass judgment on the lives it depicts *at the same time as* it challenges readers to reassess their complicity in forms of structural violence they do not see. ”

Additionally, overlapping bundles were combined. For example, *be read as a* logged 40 tokens. 24 tokens of *be read as a* were preceded by the modal *can*. As such, *be read as a*, when separated from “can,” resulted only in 16 tokens -- not quite enough for the established threshold of 20 times per million words for four-word bundles. Similarly, *can be read as* only had 15 tokens which were not followed by the determiner “a,” which prevented it from meeting four-word bundle threshold criteria on its own. With the two four-word bundles combined, however, *can be read as a* logged a frequency of 24 tokens—more than double the minimum threshold for a five-word bundle. Also, determiners at the beginning, middle, or end of bundles of the same definiteness status but varying in gender and number were combined. For example, *en el que el/en el que los/en el que la/en el que las* [in which the] was included as one entry on lexical bundle lists.

3.3.1 Structural Lexical Bundle analysis

Each LB on the list was analyzed syntactically according to a prior structural taxonomy (Cortes, 2008). Superordinate categories for analysis included noun-phrases that were followed

by noun, preposition, and verb phrases, as well as *that*-clauses, in addition to prepositional phrases that were followed by noun phrases, prepositional phrases, adverbs, *that*-clauses, and verb phrases. Quantifiers, verb phrase sequences, and *that*-clause bundles were also part of the taxonomies. Differences in structural tendencies in both languages, particularly to those occurring in LCW but not attested in other disciplines such as history (Cortes, 2008) or psychology (Lake & Cortes, 2020), were noted.

3.3.2 *Functional Lexical Bundle analysis*

An approximation of the potential communicative purpose of each bundle was determined via the analysis of co-text horizons, in line with recommended functional taxonomies (Biber et al., 2004; Tracy-Ventura et al., 2007). Table 3.13 summarizes the categories in extant taxonomies with an example of each.

Table 3.13. Functional categories of LBs with explanations of criteria and examples

Category	Description	Example
Framing Attributes (Biber et al., 2004)	Refers to the attributes of intangible/abstract or concrete/tangible phenomena	In the case of
Obligation/Directive (Biber et al., 2004)	Describes an action as a matter of obligation	It is necessary to
Place Reference (Biber et al., 2004)	Refers to concrete places	The end of the
Time-Event (Biber et al., 2004)	Refers to a specific time or event in place	At the time of
Stance Epistemic Impersonal (Biber et al., 2004)	Expresses stance/opinion toward a matter	The fact that the
Structural Only (Tracy-Ventura et al., 2007)	Syntactic building blocks	<i>En la que se</i> [in which]
Topic Elaboration / Clarification (Tracy-Ventura et al., 2007; Cortes, 2008)	Used to extend the amount of information allocated to a given topic	On the other hand

Table 3.13 Continued

Topic Introduction / Focus(Tracy-Ventura et al., 2007; Cortes, 2008)	Used to shift to a new topic	Was one of the
--	------------------------------	----------------

3.4 Communicative Purpose identification

In an earlier pilot study (Lake & Cortes, 2019), sentences in 8 research articles were categorized according to a taxonomy derived from Stage 1 of Mayring's (2014) Qualitative Content Analysis, in which analysts work independently to reduce a string of text to its essence via inductive categorization techniques. Table 3.14 shows a preliminary example of the hierarchy. Each category is discussed in greater detail in section 3.6.1.

Table 3.14. Communicative purposes hierarchy

General	Specific	Definition
Reporting	Defining	A definition of theoretical constructs or other key terms.
	Describing	A description of a person or a state of affairs on a real or imagined plane.
	Retelling	A restating of information, whether from the author's biographical details, research by other scholars, or the novel itself.
Inferring	Explaining	A presentation of information with some degree of inference not immediately obvious to the reader.
	Evaluating	Evaluations containing evidence of an attitude toward a phenomenon.
	Comparing	Can be used to establish an interplay of similarities between: Characters, Elements in the plot, Other scholars, or other works of fiction and the theoretical framework

3.4.1 Rater training

The purpose of the communicative purpose identification of bundles phase of the project was to determine whether lexical bundles appear to exert any influence on specialists in their assessing the communicative purposes of sentences in literary criticism articles. To this end, an institutional review board form (available in Appendix R) was approved for permission to request the assistance of four graduate students (two majoring in Spanish-language literature and two in English-language literature) familiar with literary criticism articles. They were recruited as second raters and briefed on the goals of this project in exchange for an honorarium, generously provided by the Honor Society of Phi Kappa Phi Love of Learning Award.

After informed consent was obtained from each participant, they completed an hour-long training on Google Hangouts using the screensharing feature to practice using the communicative purposes hierarchy shown in Table 3.14. Tables 3.15-3.20 contain abridged versions of the training materials given to the participants in English and Spanish. Anchor sentences in English and Spanish appear underneath the descriptions and guidelines for relevant linguistic features to inform their judgments. To ensure that lexical bundles would not exert any preconceived influence on the raters' judgments, the researcher chose articles from the two corpora containing the smallest volume of four-word lexical bundles:

- George, J. (2014). Suburbia's Alien Reek: Contract and Relation in John Cheever's *Bullet Park*. *Canadian Review of American Studies*, 44(3), 526-546.
- Nieto, E. F. (2014). Novela policiaca y hermenéutica: el colapso de la argumentación del detective como metáfora epistémica de *Blanco nocturno* [Detective novel and hermeneutics: the collapse of the detective's argument as epistemic metaphor of *Blanco nocturno*]. *Anales de Literatura Hispanoamericana* 43, 439-452.

The author then coded every sentence in the above articles using the communicative purposes taxonomy (Lake & Cortes, 2019). Finally, the sentences were transferred into training materials.

The worksheets given to participants are available in Appendices D and E.

Table 3.15. Defining

Comm. purpose	Defining: A definition of theoretical constructs or other key terms.
Relevant linguistic features	Authors or infrequent theoretical concepts as the sentence subject, stative/cognition/communication verbs, additional descriptive language.
English examples	While most people only superficially interact with each other, unquestionably aping one another's assumptions and subordinating themselves to an anonymous mass that Heidegger calls "the they" —"We take pleasure and enjoy ourselves as they take pleasure; we read, see, and judge about literature and art as they see and judge; likewise we shrink back from the 'great mass' as they shrink back; we find 'shocking' what they find shocking" (164)—being authentic (eigenlicht) involves interpreting the present world in a manner different from that of the general population. (590.txt)
Spanish examples	Para la escritura Roberto Bolaño tiene una definición rotunda , que toma de un verso de Nicanor Parra: Es, dice, como ordeñar una vaca para tirarle luego la leche a la cara (Bolaño 2003a: 137). (98.txt) [For writing, Roberto Bolaño has a round definition, that draws from a verse by Nicanor Parra: that is, to say, like milking a cow in order to later throw the milk onto its face.]

Table 3.16. Describing

Communicative purpose	Describing: A description of a person or a state of affairs, real or imagined.
Relevant linguistic features	Stative/cognition verbs, descriptions of habitual actions, and verbs in the present or past perfect aspect.
English examples	Eliot is proud of his job as a middle-man at a mouthwash company, he espouses the virtues of the nuclear family, and he expresses concern for his neighbours, while Nellie is a homemaker who embraces conservative morality and traditional gender roles. (590.txt)

Table 3.16 Continued

Spanish examples	Se sospecha de un robo. (98.txt) [They suspect a theft.]
	Comprender, interpretar, no es solo una instancia científica, escribe Gadamer, sino que pertenece a la experiencia humana del mundo. [Understanding, interpreting, is not only a scientific instance, writes Gadamer, but rather it belongs to the human experience of the world.]

Table 3.17. Retelling

Communicative purpose	Retelling : Refers to a simple restating of information, whether from the author's biographical details, research by other scholars, or events taking place in the novel itself.
Relevant linguistic features	It co-occurs predominantly with the historical present, but the use of the preterit (in Spanish) is an option as well. This action is similar to "describing," but uses activity and achievement verbs rather than stative or cognition verbs. When retelling examples from the novel itself, it may co-occur with or without the support of direct quotes.
English examples	When Tony goes one step further and belittles the mouthwash business, Eliot swings his club at his son. (590.txt) Eliot's actions send Tony into a deep depression, leaving him mysteriously bedridden for months. (590.txt) Tony's odd condition, combined with Eliot's realization that he could have killed his son, forces him to recognize the possibility of death. (590.txt)
Spanish examples	Se declara inocente, pero es detenido . (98.txt) [He pleads not guilty, but is arrested.]

Table 3.17 (Continued)

	En el pueblo empiezan a circular distintas versiones de lo ocurrido. (98.txt) [In the town, various versions of the events begin to circulate.]
	El detective es el que investiga, el que mete la cabeza en lo oscuro, dice ; nada que ver con una actitud naïf, ni tampoco solo lúdica. (98.txt) [The detective is the one who investigates, the one who sticks his head into the darkness, he says; nothing to do with a naive attitude, or a playful one.]

Table 3.18. Explaining

Communicative purpose	Explaining: A presentation of information with some degree of inference that would not appear immediately obvious to the reader.
Relevant linguistic features	Stative, communication, and activity verbs in dialogue with their object-collocates; logical connectives; generalizations.
English examples	<p>Central to these debates is the language of western liberalism, specifically that of social contract theory. (590.txt)</p> <p>In this arrangement, however, what one “knows” is not the authentic self but rather the performative persona dictated by the terms of the contract. (590.txt)</p> <p>Cheever illustrates Tony’s growing alterity in a scene in which he and Eliot go golfing. (590.txt)</p>
Spanish examples	<p>Bolaño usa el desierto como espacio para esas respuestas: una metáfora muy visual para la búsqueda durísima, y peligrosa, que es la literatura. (98.txt) [Bolaño uses the desert as a space for those answers: a highly visual metaphor for the grueling and dangerous journey that is literature.]</p> <p>En esta misma brecha abierta Ricardo Piglia busca su espacio. (98.txt) [Within this very gap Ricardo Piglia seeks his place.]</p> <p>Ricardo Piglia, con <i>Blanco nocturno</i> (2010), una novela policiaca vuelta del revés, cuestiona el artificio de este género para reflexionar, en un ejercicio metaliterario que queda superpuesto a la trama, sobre las claves de la interpretación. (98.txt) [Ricardo Piglia, with <i>Blanco nocturno</i> (2010), a detective novel turned backwards, questions the artifice of this genre to reflect, in a metaliterary exercise that remains superimposed to the plot over the codes for its interpretation.]</p>

Table 3.19. Evaluating

Communicative purpose	Evaluating: Evaluations which may come across as more subjective than experts would normally allow in other disciplines.
Relevant linguistic features	Impersonal generalizations; words expressing deficit e.g., “lacks” “more...than;” attributive adjectives expressing affect, i.e., “compelling;” pronominal adjectives expressing affect (a notable phenomenon in Spanish).

Table 3.19 Continued

English examples	This emphasis on the monstrous and shallow indicates that many writers cast the blame for suburbia's failures , not on the residential model, but on the actual people who live there. (590.txt)
	However, Hoover was not interested in mere shelter and certainly did not intend to increase public housing. (590.txt)
	The surrealism Cheever employs in Bullet Park undermines the romance of this authenticity, revealing it to be too singular and self-absorbed for any sustainable form of community. (590.txt)
Spanish examples	Habría que inventar un nuevo género policial, la ficción paranoica. [One would have to invent a new detective genre: paranoid fiction.]
	Es lo más sensato: la hermenéutica, como filosofía, como epistemología, atenta sobre todo al método, tiene que ser concienzudamente cauta con la articulación de sus propuestas. (98.txt) [It is the most sensible: hermeneutics, like philosophy, like epistemology, attentive above all to method, has to be painstakingly aware with the articulation of its proposals].

Table 3.20. Comparing

Communicative purpose	Comparing: Can be used to establish an interplay of similarities between: characters, elements in the plot, other scholars, other works of fiction, the theoretical framework, etc.
Relevant linguistic features	Discourse markers may help signal the presence of a comparison or contrast, but depending on the author, the reader may become tasked with identifying one sentence as similar to or different from another.
English examples	Scott Donaldson believes that suburban contracts misapply the individualist, agrarian spirit of Jefferson and Thoreau , while Milette Shamir and F.M. Coleman consider the relation between subjectivity and nature in the American transcendentalists a forerunner to suburban land development and contracts. (590.txt)
	Like his earlier and better known fictional neighbourhood, Shady Hill, Cheever's Bullet Park is a bizarre and pluralistic society inhabited by middle-class oddities. (590.txt)
	Where Nellie recreates her identity in a place she hates, Eliot has his persona disrupted in the place he loves, his home in Bullet Park. (590.txt)

Table 3.20 (Continued)

Spanish examples	<p>Con Croce, su protagonista en Blanco nocturno, Piglia se planta ante la filosofía, ante los modos prestigiados del pensamiento, ante el método, que tuvo con Hegel su apoteosis. (98.txt)</p> <p>[With Croce, his protagonist in Blanco nocturno, Piglia places himself before philosophy, before the ways of those prestiged in thought, before the method, that Hegel exhibited with his apotheosis.]</p>
	<p>Pero Piglia, al hacer de la investigación de su comisario algo rudimentario, poco profesional, más intuitivo que racional, hace con la novela policiaca lo que Cervantes hace con la novela de caballería: Replantearla desde dentro, cuestionando su propio mecanismo, hasta hacerla colapsar. (98.txt)</p> <p>[But Piglia, upon making something rudimentary from the investigation of his Commissioner, hardly professional, more intuitive than rational, makes with the detective novel what Cervantes makes with the chivalry novel: resurrect it from within, questioning its own mechanism, to the point of making it collapse.]</p>

After discussing some practice applications of the taxonomy, the participants were given a Google Sheets spreadsheet containing sentences from some literary criticism articles. Each spreadsheet row contained a sentence from five passages randomly sampled from five research articles in Spanish (a total of 10,931 words) and six randomly sampled passages from English-language research articles (consisting of 12,694 words). Figure 3.5 shows one worksheet from the English-language dataset workbook. The first column contained a sequential sentence number. The second contains a sentence in which a lexical bundle was used. In the next column, the participants were told to choose a category from the communicative purposes hierarchy. Finally, in the “Rationale” column, participants were instructed to write any salient phrases or observations that informed their judgments that led them to choose a specific communicative purpose category.

	A	B	C	D
1	#	Sentence	Communicative Purpose	Rationale
2		Before outlining ongoing discussions of compassion and critiques thereof, it is important to define compassion and to outline how the current social climate enables social actors to feel this response from the outset.	Choose...	
3		Nussbaum explains that compassion is also grounded in a sense of human flourishing as well as an understanding of the resources that must be present for a person to succeed.	Choose... Defining	
4		In the case of the United States, these factors give rise to an idealized notion of U.S. citizenship that is grounded in the logic that we all suffer in the same way and thus must deploy similar strategies to overcome hardship.	Describing Retelling Explaining Comparing Evaluating Choose...	

Figure 3.5. Screenshot of Work Area

Lexical bundles have been found to occur more frequently in Spanish than in English (Cortes, 2008), a tendency which appears to be manifest in the current dataset of 105 sentences in Spanish and 52 in English. Since the English-language dataset contained about half as many sentences and words as its Spanish-language counterpart, participants working with the English-language materials were paid \$83 for their time, whereas those working with the Spanish-language materials were granted \$167 as a consequence of the higher workload. Funds for compensation of the participants was graciously provided by the Honor Society of Phi Kappa Phi.

3.4.2 Passage selection

To boost the probability of finding lexical bundles to assess, articles were sorted by the highest percentage of words constituted by lexical bundles. Table 3.21 shows the number of

words under the umbrella of four-word bundles in the articles chosen, as well as the percentage of the overall word count constituted by lexical bundles from each article.

Table 3.21. Mean word count constituted by LBs per article (SD) in words and % of word count

Data	Spanish	Words	English	Words
Mean	3.61%	202.5	1.42%	91.1
SD	1.24%	99.4	0.60%	40.7

While initially three-word lexical bundles were included in the present analysis, for practical and financial considerations of funding available, four-word bundles were chosen as the focal point of analysis for participants. Additionally, some articles did not have section headings, making it impractical to choose a given segment of the article to select. Such articles were omitted from consideration, leaving the final set displayed in Table 3.22.

Table 3.22. Word count, Article reference number, Tokens of LBs, and Research article section for texts chosen in the present study

English	Article	4 word Formulaicity	Passage Word Count	Bundle Tokens	Section Title
	45	1.60%	1,658		Bare Life, Guilt, and the 9 Sovereign Subject
	314	1.71%	2,532		Professor and James Reese 8 Europe as New Negroes
	318	1.68%	2,315		Postcolonial Violence and the Hard Boiled Tradition: The Case 4 of Bouaffesse
	375	1.72%	4,904	16	The Space of the Refugee: Spatial Displacement and the Loss of Home
	471	1.78%	2,388		7 Compassion and the Intimate Public Sphere
	562	2.77%	855		8 Introduction
Spanish	12	0.74%	1,048		2 Introducción (Introduction)

Table 3.22 (Continued)

78	0.73%	1,963	III. Figura europea para utopía latinoamericana: el nómade inmóvil [III. European figures for the Latin American utopia: the stationary nomad]
155	1.13%	4302	EL ÁRBOL DE LA CIENCIA : NOVELA AUTOBIOGRÁFICA [The tree of knowledge: Autobiographical Knowledge]
156	0.76%	4,431	3. INTERTEXTUALIDADES LITERARIAS EN RETORNO DE UN CRUZADO [Literary intertextualities in Return of a Crusader]
180	0.76%	1,078	4. Exclusiones, sujetos y subjetividades [Exclusion, 8 subjects, and subjectivities]

Additionally, while one way to choose the passages would have been to choose the passage containing the most lexical bundles, the researcher chose to randomly sample sections, so as to allow for a wider representation of the kinds of sections one might encounter in LCAs. Tables 3.23 and 3.24 list the bundles that occurred in the chosen passages, organized by their frequencies of occurrence.

Table 3.23. LBs by tokens found in the above passages (English)

Tokens	Bundles
1	as a form of, as a result of, as one of the, as part of a, as part of the, as well as a, as well as the, at the same time, can be read as, from the perspective of a, in a way that, in the context of, in the name of, it is important to, that there is a, the end of the
2	at the heart of, at the same time, can be read as a, in relation to the, in the case of, in the form of, of the novel is, on the level of the, the fact that the, the role of the
3	on the one hand, on the other hand, the figure of the, the way in which
4	in the United States

Table 3.24. LBs by tokens found in the above passages (Spanish)

Tokens Bundle	
1	<p>a la vez que, a partir de sus, a partir de una, a pesar de la, a pesar de que, a sí mismo como, a través de las, a través de los, a través de su, a través de un, al final de la, cada uno de los, de la casa de, de la historia que, de la muerte de, de la novela como, de la novela es, de la novela y, de la obra de, de la obra del, de la vida de, de la vida del, de las novelas de, de los personajes que, de una serie de, del protagonista de la novela, desde el punto de vista, desde la perspectiva del, el punto de vista, en el caso de la, en el caso del, en el momento de, en el sentido de, en el siglo de las luces, en el siglo xix, en la medida en que, en la novela de, en la novela los, en la que se, en la vida de, en lo que se refiere a, en los que se, en relación con la, en torno a las, la construcción de un, la forma en que, la historia de la, la lectura de la novela y, la mayoría de los, la posibilidad de un, la posibilidad de una, la problemática de la, la vida y la, más allá de la, no deja de ser, que se refiere a, se convierte en un, se manifiesta en las, se trata de la, se trata de un, y por lo tanto</p>
	<p>[at the (same) time that, from her/his/its, from a, in spite of the, despite, to herself/himself/itself as, through the (m), through the (f), through her/his/its, through a, at the end of the, each one of the, of/from the house of, of the history that, of the death of, of the novel as, of the novel is, of the novel and, of the work(s) of, of the work(s) of, of the life of, of the life of (the), of the novels of, of the characters that, of a series of, of the protagonist of the novel, from the point of view, from the perspective of (the), the point of view, in the case of the (f), in the case of (m), at the time of, in the sense of, in the century of enlightenment, in the nineteenth century, insofar as, in the novel of, in the novel the (m. pl.), in which + [deagentivization particle], in the life of, in what refers to, in which + [deagentivization particle], in relation to the, concerning the, the construction of a, the form/manner/way in which, the history of the, the reading of the novel and, the majority of the, the possibility of a, the possibility of a (f), the problematic of the, the life and the, beyond the (f), does not cease to be, which refers to, becomes a, manifests itself in, is about, is about (f), and therefore]</p>
2	<p>a pesar de su, de la historia de, de la historia y, de la novela en, de la novela que, el hecho de que, en el que se, en la novela la, en la novela que, en las que se, la historia y la, parte de la novela</p>
	<p>[despite its, from the history/story of, from the history/story and, of the novel in, of the novel that, the fact that, in which, in the novel the, in the novel that, in which, the history/story and the, part of the novel]</p>
3	<p>a partir de la, el protagonista de la novela, en la obra de</p>
	<p>[from the, the protagonist of the novel, in the work(s) of]</p>
4	<p>el momento en que</p>
	<p>[the moment in which]</p>
7	<p>en el caso de</p>
	<p>[in the case of]</p>

3.4.3 *Lexical Bundle Communicative purpose analysis*

The researcher checked the completed spreadsheets for any discrepancies between the two raters. The researcher also completed each worksheet in English and Spanish using his own familiarity with the Communicative Purposes Hierarchy. If discrepancies of more than 2 degrees of separation on the scale were detected between the researcher and both participants per language group, the rationale cell was consulted for clarification. Afterwards, the second raters' responses were analyzed through the lens of the QCA-derived taxonomy (Lake & Cortes, 2019) in the chosen sentences.

The researcher tallied up agreement points between himself and the other two participants using the following values:

Table 3.25. Agreement score calculation criteria

Category	Conditions	Points
Full	All both participants agree with the researcher	1.00
3/4	The researcher agreed with at least one of the participants	0.75
Half	The researcher agreed with both raters that the sentence was somewhere on the “reporting” or “inferring” end of the hierarchy	0.50
1/4	The researcher agreed with one rater that the sentence was somewhere on the “reporting” or “inferring” end of the hierarchy	0.25
None	No exact or adjacent agreement between me and the other two raters	0

An additional measure of bundle saliency was calculated based on the following criteria: each participant and the Experimenter who mentioned a bundle in their response by name contributed a score of +0.33 to the bundle saliency score for each sentence in the dataset. This process is summarized with the following formula:

$$\begin{aligned} & ((\text{Raters who mentioned the bundle in their Rationale} * 0.33) + \\ & (\text{Raters who alluded to the content of a bundle} * 0.165)) \\ & = \underline{\text{Bundle saliency score}}. \end{aligned}$$

To illustrate, take the following sentence:

- (1) Before outlining ongoing discussions of compassion and critiques thereof, **it is important to** define compassion and to outline how the current social climate enables social actors to feel this response from the outset. (471.txt)

Participant A-EN (Alias “Kay”) wrote as their rationale: “...is important to define...”. While only 3 of the four words in the bundle were included, for the purposes of this project, it was coded as full saliency in the judgment. Thus, 0.33 was added to the bundle saliency score. The Experimenter, as a rater, wrote “Declaring something important” which was also deemed as enough of a connection to saliency, netting another 0.33. Participant B-EN (Alias “Robin”), however, focused instead on the verb “define,” writing the following: “I want to say this is Defining because the word ‘define’ is in the sentence. Though, the essayist technically did not yet define ‘compassion’ here, so I might be wrong,” resulting in a score of 0 in saliency drawn from Robin. Adding the scores for the researcher and Kay together, however, the bundle saliency score for the above sentence was 0.667.

Similarly, take the example sentence from the Spanish dataset: “ése es **el momento en que** se inicia la acción en El árbol de la ciencia” (155.txt) [That is the moment in which *The Tree of Knowledge’s* action begins]. Participant A-ES (alias “Jaime”) said, of the above sentence, that “ése es el momento...” presenta información que cambia el eje de la historia. [‘That is the moment....’ he presents information that changes the axis of the story.]. Participant B-ES (Alias “Adán”) wrote: “Summarizig[sic] events, marks the beginning of something.” Finally, the Researcher wrote “Describing timeline.” All of the above make a direct mention either to the

word moment or to the closely semantically related concept of time, earning a saliency score of 1.00.

If the Researcher saw potential for an indirect allusion to the bundle in the rater's responses, +165 was added to the saliency score for each sentence. For instance, take the case of the sentence "Con excepción del ambiente familiar, donde el autor introduce algunos cambios con respecto a su vida real, el resto de pasos que da **el protagonista de la novela** son los mismos que dio él mismo" [With the exception of the family environment, where the author introduces some changes with respect to his real life, the rest of the steps taken by the protagonist of the novel are the same ones that he himself took.]. Jaime wrote that "Estos son detalles **biográficos del autor transpuestos a la novela**" [These are **biographical details from the author transposed onto the novel**]. Adán based their decision on the following rationale: "reference to actions taken by a character." The researcher wrote "Mismo [same] -- comparison of the two characters." These three rationale statements were all interpreted as indirect. There are multiple possible human entities in the sentence to which they referred. Thus, it is not certain whether the focal point of the sentence was the protagonist or the novel's author. In cases such as this, a conservative interpretation of the data was chosen, yielding the net bundle saliency for the three respondents in this sentence a value of 0.495.

3.5 Conclusion

In summary, this chapter has reported on the methodological approaches used in the present study. Chapter 4, addressing Research Question 1, will offer the results of the Multidimensional analysis in English and Spanish. Chapter 5 will answer Research Question #2:

the kinds of equivalent and language-specific bundles occurring in LCW. Finally, Chapter 6 will address Research Questions 3 and 4: the communicative purposes of lexical bundles and the range of communicative purposes identifiable in LCW.

4 REGISTER ANALYSIS AND MULTIDIMENSIONAL ANALYSIS OF ENGLISH AND SPANISH CORPORA

Research Question 1: Multidimensional Analysis of LCW in English and Spanish

The first research question seeks to determine the characteristics of LCW in English and Spanish via multidimensional analysis. Accordingly, this chapter examines LCW using register analysis and multidimensional analysis frameworks. Section 4.1 offers remarks on LCW in both languages using Gray's (2015) framework for academic journal articles. Section 4.2 contains the multidimensional analysis of the English-language LCW corpus. Section 4.3 holds the multidimensional analysis of the Spanish-language LCW corpus. Finally, Section 4.4 offers some remarks on cross-linguistic similarities and differences revealed in the analysis.

4.1 Register Analysis of Literary Criticism in English and Spanish

In accordance with recent practices for using MDAs to study specific academic disciplines, the present section begins with the results of the register analysis whose framework (Gray, 2015) was described in Section 3.1.1. It compares and contrasts the findings in English and Spanish with those reported by Gray (2015) for journal articles in other disciplines. Table 4.1 contains the results of the English-corpus and Spanish-corpus register analyses using Gray's (2015) situational framework for academic journals.

Table 4.1. Framework for describing the situational characteristics of academic journal registers

General	Specific	Metric	English	%	Spanish	%
1. Participants	A. Writer	1 (single)	149	98%	169	97%
		2-4 (small group)	3	2%	5	3%
	A. Length	Page length	17.6		16.5	
		Standard deviation of page length	5.9		6.1	
		Range	7-31		5-38	
		None	52	34%	58	33%
		Numbered	13	9%	42	24%
	B. Headings	Un-numbered	87	57%	74	43%
	C. Use of Abstracts	Yes	152		174	
	D. Visual Elements	None	147	97%	166	95%
		Figures	5	3%	8	5%
		Introduction only	7	5%	16	9%
		Conclusion only	12	8%	15	9%
		Introduction & Conclusion	23	15%	8	5%
	2. Textual Layout & Organization	E. Sections / Organization	Variable section headings/names	100	66%	116
3. Setting	A. Nature of Journal	Generalist	152	100%	174	100%
4. Subject / Topic	A. General Topic of the Discipline	Varied (based on an informal survey of a range of sources)	152	100%	174	100%
5. Purpose	A. General Academic Purpose	Varied (based on initial register survey)	152	100%	174	100%
	A. Presence of Observed Data	Yes	152	100%	174	100%
	B. Use of Numerical Data	No	152	100%	174	100%
6. Nature of Data or Evidence	C. Primary Presentation of Evidence	Extensive prose	152	100%	174	100%

Table 4.1 (Continued)

	D. Object of Study Varied (based on trends in the corpus)		152	100%	174	100%
	A. General Method Type	Observational	152	100%	174	100%
7.	B. Statistical Techniques	n/a	152	100%	174	100%
Methodology	A. Explicitness of Purpose	Direct statement	141	93%	157	90%
		Indirect	11	7%	17	10%
	B. Explicitness of Research Questions	No discernible statement	152	100%	174	100%
	C. Explicitness of Citations	Within text and footnotes	27	18%	15	9%
		Footnotes only	125	82%	159	91%
8.	D. Explanation of Evidence	Extensive Mention	152	100%	174	100%
Explicitness of Research Design	E. Explanation of Procedures	No discernible statement	152	100%	174	100%

The following subsections contain in-depth remarks on each of the eight major areas of the register analysis framework.

4.1.1 Participants

First, whereas in the sciences it is more common to have multiple authors (Larivière et al., 2006), as nearly all quantitative biology and physics articles in Gray's corpus did (4% single-authored in both corpora), articles in literary criticism in both languages (EN = 98%; ES = 97%) were single-authored -- an inverted, mirror image. Larivière et al. (2006) describe collaboration in humanities writing as a "marginal phenomenon" (p. 13). A possible explanation may come in

the form of the methods used, articulated by Thieme (2017) as “close reading” conducted by one person, whereas in quantitative disciplines, one sometimes encounters a need for multiple experts to tackle larger-scale problems based on larger datasets (Iglič et al., 2017). Indeed, Given & Willson (2015) offer that collaboration may have arisen more quickly in the sciences as a consequence of needing to use different tools for different facets of experiments, whereas a humanities article is typically the product of a single individual’s perspective on a given expression of art.

4.1.2 Textual layout and organization

Table 4.2 compares mean page numbers for LCW in English and Spanish to other disciplines in Gray’s (2015) corpus. With regard to page length, there are two possibilities worth noting. Gray’s chart does not report words per page. Even so, it is interesting to note that literary criticism articles, although higher than quantitative physics (Mean = 13.4, SD = 7.3), still have a lower average page length (English: Mean = 17.6, SD = 5.9; Spanish: Mean = 16.5, SD = 6.1) than that of other qualitative disciplines in Gray’s corpus.

Table 4.2. Comparison of LCW articles and other disciplines (Gray, 2015) by page length.

	English LCW	Spanish LCW	Physics (Quant)	Applied Linguistics (Qual)	History (Qual)
Mean	17.6	16.5	13.4	20.8	24.3
SD	5.9	6.1	7.3	4.9	6.7
Range	7-31	5-38	5-32	14-32	8-42

One contributing factor to this detail may be that the present study deals specifically with literary criticism articles analyzing just one novel. It is possible that the page numbers for LCAs could approach those of other qualitative disciplines if comparisons between two novels, an

approach that has been promoted in literary studies more recently in the field (de Zepetenek, 2007), were selected.

With regard to section headings, the majority of articles in Gray's corpora used either numbered or un-numbered headings. 67% of history articles, however, did not use subject headings. In the present study, about one third of articles (EN = 34%; ES 33%) did not use subject headings. While analyzing the specifics of subject headings is beyond the scope of this dissertation, a partial replication of Lin & Evans (2012), who analyzed section heading patterns in a wide range of research articles, could yield more insight into the macrostructure of LCAs. LCW also closely resembled applied linguistics (qualitative and quantitative) and history concerning its numbered headings to un-numbered headings ratio. However, Spanish articles favored numbered headings more (24%) than did English articles (8%).

As for the use of abstracts, the present corpus is not representative of all literary journals. Some journals were excluded from the present sample due to their high quantities of articles published without abstracts. Nonetheless, since abstract writing skills are argued to be a skill of high importance for breaking into other discourse communities, for example, in applied linguistics conference presentations (Payant & Hardy, 2016), abstracts in LCW were deemed to be of importance in the present study.

Additionally, unlike LCW in both languages, the majority of other disciplines rely heavily on some combination of Tables, Figures, and/or Equations. LCW articles that did avail themselves of these techniques usually did so with images of works of art relevant to the argument. In the avenue of visual elements, the discipline which LCW most heavily resembles is theoretical philosophy, which also had a 3% usage of Figures.

A central component of LCW is argumentation based on quotes from the novels studied, and sometimes a philosophical, theoretical framework. Specifically, Tötösy de Zepetnek (2007) states in more recent humanities writing that “no matter what cultural product a scholar analyses or investigates, a theoretical framework and methodology must be part and parcel of the work” (p. 59). Examples 1 and 2 are examples of theoretical frameworks represented in the corpora:

- (1) Empleando las teorías actuales sobre la construcción del género y de la sexualidad, sobre todo las de Judith Butler y de la performatividad queer, este ensayo analiza cómo los personajes de Schoo sirven como ‘estudios de caso’ de la deconstrucción de los conceptos tradicionales sobre el género y la sexualidad. (18.txt)
[Utilizing current theories on gender construction and sexuality, above all, those of Judith Butler and queer performativity, this essay analyzes how Schoo’s characters serve as ‘case studies’ of the deconstruction of traditional concepts of gender and sexuality.]
- (2) It is in this context that Mabasa establishes a link with oral traditions which function as the vital nourishing supplement to his creative act. Such a creative modality is not only in sync with the creative demands of Afrocentricity, a theory that emphasizes the placement of ‘African ideals at the centre of any analysis that involves African culture and behavior’ (Asante, 1998:2), but is also ideologically and pedagogically empowering as it elevates orature to a position where, ‘it is the incontestable reservoir of the values, sensibilities, esthetics, and achievements of traditional African thought and imagination .’ (277.txt)

Additionally, Viera Echevarría (2014) found that argumentation strategies were present in the introduction, body, and conclusion phases of Spanish-language literary criticism conference presentations in Spanish. Thus, ubiquitous argumentation strategies, coupled with discussion of philosophical-theoretical constructs, may contribute to LCW’s resembling philosophy in this area.

As for standardized sectional headings such as the IMRD pattern favored in empirical research articles, no literary criticism used an IMRD organizational macrostructure. In the

English LCW corpus, however, 28% of articles explicitly marked either an introduction, conclusion section or both in the paper. Spanish did so in 23% of the articles.

However, despite the lack of marking some sections as such, there is some evidence to suggest that LCW utilizes some communicative purposes present in IMRD articles, albeit in ways that would not be obvious to most readers. For example, Viera Echevarría (2014) adds that, at least in literature conference presentations, the introduction may state the purpose.

Conclusions may opt to restate the thesis or suggest further work in the area, or else, use an epilogue, which may or may not state the thesis but attempt to leave a strong impression on the listener. Additionally, Thieme (2017) proposes that some literary criticism articles may, to a mild extent, use noun phrases to specify a paper's methodological approach. Such examples include "analysis of selected literary depictions of residential school survivors that focus on gender segregation and the shaming of the body", "a more nuanced consideration of the complexity and heterogeneity of Asian and North American subjectivities", and "my analysis of the novel in terms of assemblage theory" (p. 101). The extent to which this phenomenon takes place warrants larger-scale scholarly attention.

4.1.3 Setting

Similar also to philosophy, in Gray's study, the majority of articles in this area came from generalist journals. We also note that the articles in our corpora do fall under the umbrella of certain regional specifications. *ISTMO: Revista virtual de estudios literarios y culturales centroamericanos*, for instance, may study novels in myriad genres. Yet one similarity between most articles published in *ISTMO* is that the texts are analysis of Central American authors.

Similarly, *MELUS* stands for Multi-Ethnic Literature of the U.S. Thus, it publishes analyses of works of literature by writers of various ethnicities who have spent time in the U.S. prior to composing their novels. However, the novels studied in these areas can vary widely by time period and characteristics of the characters involved. We thus opt to classify them as generalist journals for the purposes of the present study.

4.1.4 Subject/Topic

Articles in both corpora aligned in topic. Articles in this discipline do report on an analysis of observed data, albeit through argumentation from observations on a single novel.

4.1.5 Purpose

A unifying principle in the purpose of these articles appears to be interpretations of specific novels as potential focal points from which to wrestle with larger questions of the human experience. To illustrate, the journal *Literatura Mexicana*, for instance, describes its purpose as being to disseminate knowledge of Mexican literature, as well as authors, and genres, and their ideas, composed in Mexico from the 16th century to the present. Similarly, MELUS aims “to publish essays that advance ongoing critical conversations about the theoretical, historical, literary, and cultural contexts of multi-ethnic literature, film, and other kinds of texts” (MELUS: Submissions).

4.1.6 Nature of data or evidence

The presence of observed data separates literary criticism from philosophy due to its need to include textual examples from multiple portions of the same novel to advance the author's argument. However, there is no numerical data. In relation to the primary presentation of evidence, however, LCW does resemble theoretical philosophy, qualitative history, qualitative political science, and qualitative applied linguistics. It is interesting that it aligns with the theoretical aspect of philosophy, but not theoretical physics, which instead uses mathematical formulae. It is also observational like the other qualitative disciplines in Gray's corpus. As for the object of study, those in this corpus were one novel from the 20th century onward per article.

4.1.7 Methodology

As previously discussed, Thieme (2017) describes the methods of the literary scholar as a "close reading" of a text, sometimes putting it in dialogue with other texts. These might be called observational, if only as the observations of the researcher put in dialogue with other scholars' observations.

4.1.8 Explicitness of research design

Concerning the explicitness of purpose, the majority of articles state the purpose of the study (93% in English and 90% in Spanish). Tankó (2017) remarked that only 85.2% (N = 135) of literary criticism included an observable statement of purpose. Examples 3 and 4 contain examples of statements identified as purpose statements in the corpora:

- (3) En el presente trabajo me interesa considerar los modos en que Ricardo Piglia se ha constituido como escritor que lee y ha hecho de esta operación una clave de

sus materiales, no sólo en la crítica argentina, sino desde su doble localización como profesor en Estados Unidos y Buenos Aires. (177.txt)

[In this paper I am interested in considering the ways in which Ricardo Piglia has constituted himself as a writer who reads and has made this operation a key to his materials, not only in Argentine criticism, but from his double location as a teacher in the United States and Buenos Aires.]

- (4) This paper reads Kim Barry Brunhuber's novel *Kameleon Man* as an important exploration of different heterotopian spaces offering a reflection of our contemporary society in terms of the production and consumption of culture, racialization and identity. (531.txt)

Based on the register analysis criteria, these data remain comparable with the numbers for explicit purpose statement delineated in abstracts (Tankó, 2017) and introductions in other disciplines (Gray, 2015). However, one reason for this difference may be that Tankó's sample looked only at abstracts, whereas Gray's (2015) register analysis allows the researcher to search for a purpose statement within the first pages.

Concerning research questions, no article in LCW used them, opting instead for purpose statements. In this regard, LCW most closely resembles theoretical philosophy, theoretical and quantitative physics, as well as qualitative history.

Gray (2015) operationalizes "explicit" citations as ones consisting of an author's name and date included in the body of the article (e.g., "the numbers for explicit purpose statement delineated in abstracts (Tankó, 2017)" as well as an author's name used in conjunction with a reporting verb (e.g., "Gray (2015) operationalizes"). Articles with less explicit citations make a fuller use of numbered footnotes that link to bibliographical citations. As Afros & Schryer (2009) attest, however, common purposes of the footnote in LCW include self-mentions of literary critics, communicative bibliographical details about a work cited, comments deemed tangential but still worthy of inclusion in the study, as well as supplemental examples and quotations. In this regard, LCW differs from other disciplines in its tendency toward both

“explicit” citations in the body of the article and explicit citations embedded in footnote discourse segments ranging in length from a sentence to several paragraphs. Example 2 in Section 4.1.2, for example, contains the following explicit citation: “Afrocentricity, a theory that emphasizes the placement of ‘African ideals at the centre of any analysis that involves African culture and behavior’ (Asante, 1998:2)” LCW most closely resembles quantitative and qualitative applied linguistics’ levels of explicit citations.

Finally, inasmuch as an explanation of evidence is concerned, Gray (2015) writes that “the distinction is between articles which contain developed descriptions of data and methods, and those which primarily omit explicit information about these aspects of research” (p. 65). As Thieme (2017) pointed out, LCAs that describe methods or evidence in any detail comparable to that espoused in other disciplines is a rarity. The two disciplines in Gray’s (2015) study that mirror the tendency of not explicitly describing evidence are qualitative history and theoretical philosophy, reflecting the need of interlacing philosophical explanations for observable narrative phenomena in novels.

4.2 RQ 1: Multidimensional Analysis of LCW in English and Spanish

With the register analysis completed, Research Question 1 is explored in Section 4.2 and its subsections, which discuss the results of the MDA of linguistic features performed on the English-language corpus. After an initial discussion of the results, some salient features of the six dimensions are discussed in relation to knowledge about writing in other disciplines. Discussions of each dimension include passages from text files with low, medium, and high, scores on said dimension. This decision was made following the example of Biber (1988), who discusses the

implications of not only medium scores on dimensional scales in written and spoken texts, but also maximum and minimum values. In the present study, this practice is complemented with the presence of qualitative data to inform a discussion of the texts from which the scores are generated.

4.2.1 English MDA results

The purpose of the MDA performed in the proposed study is to reveal what varieties of language LCW most closely resembles. The present chapter discusses findings along six dimensions, which are part and parcel of many MDA studies (e.g., Biber, 1988, 1995; Biber & Finegan, 1994; Biber & Conrad, 2001; Connor & Upton, 2003; Geisler, 2003; Al-Surmi, 2012; Pérez-Paredes & Sánchez-Tornel, 2019).

By MDA, we refer to the dimension scores generated for each text file described in Chapter 3, using Biber's (1988) factor scores obtained for a wide variety of texts as referential criteria. These scores are then used to plot on a graph which other registers literary criticism resembles. It should be reiterated that "positive" and "negative" loadings are not "value judgments" on the results. Rather, they simply refer to the extent to which a number sits above (positive) or below (negative) the mean scores reported for the conglomeration of linguistic features reported by Biber (1988) based on measurements of each feature taken from approximately 1.5 million words of English text.

Table 4.3 contains the MDA results.

Table 4.3. English MDA results

Dimension	1	2	3	4	5	6
(Mean (SD))	-18.2 (4.0)	-2.3 (0.9)	8.4 (2.2)	-4.0 (0.9)	3.8 (1.8)	1.4(1.8)

The scores show, for example, 68% of LCWs sit somewhere between -22.2 and -14.2 standard deviations below the mean scores obtained for a general distribution of language features with a balanced level of interpersonal and informational features. The scores obtained for Dimension 6, conversely, show that 68% of text files in the corpus possessed between -0.4 and 3.2 standard deviations above the mean for levels of demonstrative determiners and noun & verb complement clauses. In Sections 4.2.1-4.2.6, we discuss in greater detail how these scores reflect LCW's unique lexicogrammatical profile in relation to other registers.

Table 4.4 contains the interfactor correlations for the dimensional density scores in the Spanish LCAs. The data were not normally distributed. Thus, a nonparametric Spearman's Rho (cf. Larson-Hall, 2010), rather than a Pearson's R, was calculated for the correlations between each factor.

Table 4.4. English MDA interfactor correlations

Dimensions	1	2	3	4	5	6
1	1.00					
2	0.40***	1.00				
3	-0.37***	-0.17*	1.00			
4	0.64***	0.20*	-0.31***	1.00		
5	n.s.	n.s.	n.s.	0.30***	1.00	
6	0.40***	0.31***	n.s.	0.34***	n.s.	1.00

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Calculating inter-factor correlations is important because redundant variables in statistical analysis leave open a greater potential for misleading results. One strategy recognized as minimizing its effects is to inspect correlations between variables to ensure that no two in a given model have significant correlation coefficients with absolute values greater than 0.7 (Slinker &

Glantz, 1985). Table 4.4 reveals that all interfactor correlations were below 0.7 in the present study.

Finally, we note that in our presentations of qualitative examples in discussions of each dimension score, manual inspections of text files with low, medium, and high dimensional scores on each dimension were carried out using regular expressions in Geany (Tröger et al., 2019), an open-source text editor for Linux operating systems, to highlight passages with high, medium, and low densities of dimensional features in an attempt to, qualitatively, present a relatively balanced view of what the scores translate to in article content. This decision was made in order to provide qualitative examples informed by Biber’s methodological protocol to examine the “internal coherence of a genre category – that is, the range of variation possible within a genre” (p. 171). Figure 4.1 shows Geany working as a rudimentary detector of positive loading Dimension 1 features in a specially prepared version of one of the tagged research articles in the corpus.

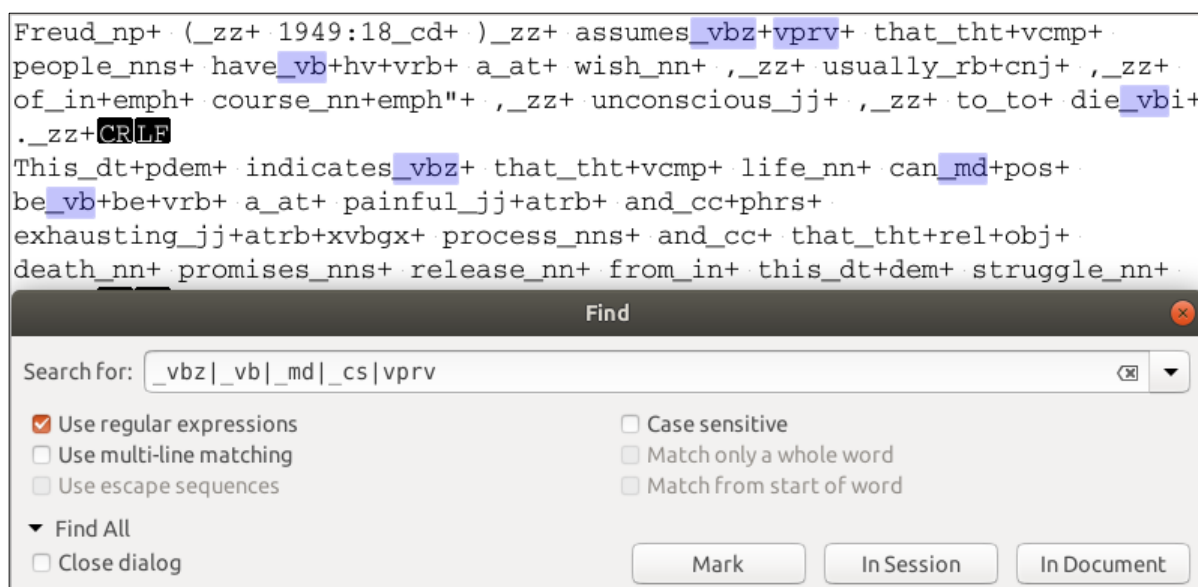


Figure 4.1. Geany’s search box and regular expressions detecting concentrations of positive-loading Dimension 1 features in proximity.

To obtain the visualization in Figure 4.1, the tags associated with dimensional features were put into the search box. The tag “vbz”, for instance, is associated with present tense, third person verbs. “Vb” can sometimes result in Geany locating first and second person present tense verbs. However, such cases as “die_vbi”, indicating that it is an infinitive, warrants manual inspection. Each tag in the search query is separated by the pipe (“|”) character, which translates in the regular expression query language to “or”. Thus, we can think of the query in the search box of Figure 4.1 as “search for present tense third person verbs OR verbs that can sometimes be present tense OR modals OR subordinating conjunctions OR private verbs”. Then, a click of the “Mark” button in the search bar causes all results to stand out in a charming shade of periwinkle.

4.2.2 Dimension 1: Involved vs. Informational Production

Dimension 1, about which there is the most to discuss, distinguishes between “Involved” and “Informational” production circumstances as the texts are produced. Biber (1988) describes involved texts as ones generated within the confines of unplanned time frames for production which are often “affective, interactional, and generalized” (p. 107) in nature. Based on the linguistic features associated with these dimensions, Biber (1988) classifies telephone conversations as the most involved. Informational texts, on the other hand, are generally composed in circumstances with a richness of planning time that results in exactness and dense quantities of information condensed into fewer words. Biber (1988) found that official documents were the most informational in nature.

Dimension 1 scores for literary criticism articles were (Mean = -18.2, SD = 4.0), a score situated near the levels of academic prose (Mean = -14.9, SD = (6.0)., and official documents (Mean = -18.1, SD = 4.8). Figure 4.2 shows mean scores for LCAs in relation to other texts.

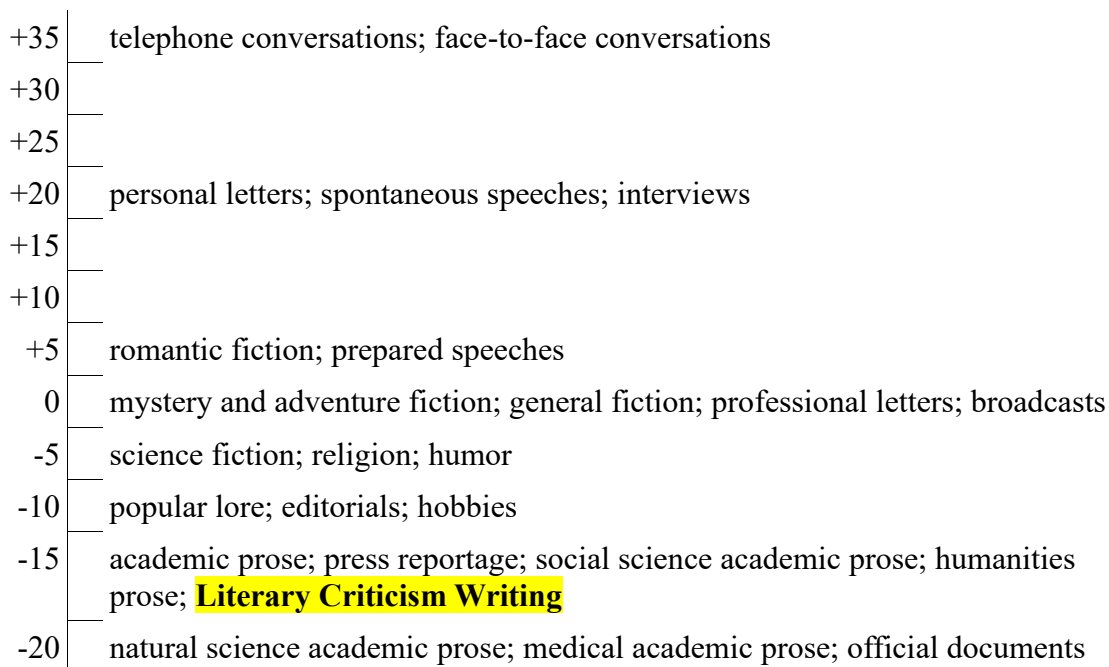


Figure 4.2. LCW articles in relation to other text varieties on Dimension 1 in English.

Documents with negative scores on Dimension 1 exhibit high volumes of attributive adjectives, prepositions, and nouns. Documents yielding negative Dimension 1 scores may also contain few present tense verbs or coordinating conjunctions per sentence. For the purposes of illustrating negative-loading Dimension 1 prose, Example 5 displays a passage from the lowest-scoring Dimension 1 article (-29.1) in the corpus. 28.5% of the words in the passage are nouns (underlined in Example 5) while 11.9% of its lexical items are attributive adjectives (bolded). Both these lexical categories are negative-loading features on Dimension 1. Present tense verbs and the verb *to be*, which function as positive loading features on Dimension 1, are highlighted.

The character bi-gram "<>" denotes a place where direct quotes were removed during the corpus cleaning phase.

- (5) These metaphors **are** a **key example** of [what Sharrad **describes**] as Wendt's "**fictive indirectness**": narrative which **weaves** out of the text and **folds** back into itself in order to evoke **intersectional postcolonial politics** and **complex characterization** through **dense networks** of suggestion (2002, 110). This technique **is** critical to Wendt's approach to decolonizing **Pacific histories**, a methodology which **recognizes** a fluid boundary between history and fiction, thus taking up **imaginative narrative** as a **critical tool** through which to explore **psychological** and **experiential realism** (Wendt 1987). Crucially, interpreting "**fictive indirectness**" in this scene **relies** on the **spiraling structure** central to many **Pacific indigenous rhetorical traditions**: the lithograph's meaning **emerges** through the protagonist's iterative physical and **imaginative returns** to this **ceremonial rehistorying** (Wendt 1992, 30–31, 36, 242). Each new <> **introduces** another layer of interpretation until he **understands** the **total interconnection** of his own narrative of loss with a legacy of entangled violences masked by the discourse of technological utopia. He **remembers** that <> (35), and the knowledge culminates in a **ritual burial** of the lithograph in which imagery of heliotropic violence – <> (242) – **coheres** his **familial memories** of reordinarisation and violation with <> (242). (en_545.txt)

Table 4.5 displays a distribution of positive and negative loading features in the above passage.

Table 4.5. Raw and % of frequency counts of linguistic features in Figure 4.2.

Negative loading features:	nouns	55	28.5%
	attributive adjectives	23	11.9%
	prepositions	27	14.0%
	Sum	54.4%	
Positive loading features:	The verb to be	1	0.5%
	Present tense verbs	9	4.7%
	And as a coordinating conjunction	7	3.6%
	Sum	8.8%	

With 54.4% of its text consisting of negative loading features, while 8.8% of its prose is woven from positive loading features, it becomes easier to understand how the above passage contributes to the document's strongly negative Dimension 1 loading scores. Tucker (2014) describes attributive adjectives as a feature strongly characteristic of art history writing. These

remarks may also hold relevance for the centrality of attributive adjectives in LCW, another humanities discipline which analyzes and describes works of art. However, Tucker asserted that the purpose of such adjectives was aesthetic in nature. While some arguably poetic phrases appear in Example 5, such as “heliotropic violence” and “spiraling structure”, attributive adjectives also appear to serve in more technical, scholarly capacities, as with the noun phrases “postcolonial politics”, “fictive indirectness”, and “experiential realism”. Biber & Gray (2015) suggest that one distinguishing feature of humanities writing, vis-à-vis scientific writing, is longer sentences that, although still elusive in meaning to many readers, may be more feasible for many readers to parse as said nouns are often specialized operationalizations of common nouns. Evidence for nouns following this tendency in Example 5 are “politics”, “returns”, and “violences”.

In contrast to Example 5, Example 6 has fewer attributive adjectives, and more positive loading features, such as present tense verbs, the verb to be, and a demonstrative pronoun.

Example 6 comes from a text with a medium Dimension 1 score.

- (6) Caruth’s analysis of post-traumatic stress syndrome **explains**, in part, Fourth Jane’s sense of paralysis as her sense of identity **falls** apart: <> (24). The fact that Fourth Jane’s conflicting identities **consume** and <> her **makes it seem** like no small wonder that she **wants** to go to a familiar, comforting place to heal herself. Since she **has** spent so much of her life in exile, **it is** not surprising that she **exiles** herself in order to heal, travelling to France to get away from the American surroundings that **contain** so many painful memories. In her exile, Fourth Jane also **seeks** out the familiar, comforting, hospital imagery of her childhood games and memories. When she **arrives** in France, she **looks** for an apartment <> (11). This image of a stark, clean white place **evokes** the idea of a hospital; somewhere where Fourth Jane **can** **rest** and recuperate. This image of a hospital-like environment in exile also **echoes** the comfort of her childhood before the family left China, where her father was a doctor, for Europe and America, where the family **was** plunged into new and unfamiliar surroundings. In this way, Fourth Jane’s exile **becomes** a synthesis of the familiar images of medicine and travel, and the unfamiliar surroundings of a place that is removed from her old

relationships, her family, and the accompanying pressures of these ties. (en-466.txt)

Percentages of positive and negative loading feature in Example 6 are shown in Table 4.6.

Table 4.6. Grammatical features % of passage in Example 6

Negative loading features:	nouns	65	28.8%
	attributive adjectives	15	6.6%
	prepositions	30	13.3%
	Sum	48.7%	
Positive loading features:	The verb to be	4	1.8%
	Present tense verbs	14	6.2%
	And as a coordinating conjunction	5	2.2%
	Subordinating Causal Conjunctions	3	1.3%
	can	1	0.4%
	Pronoun it	2	0.9%
	who clauses	4	1.8%
	Sum	14.6%	

Thus, although negative loading features can continue to predominate LCW passages, even relatively small increases in positive loading dimension features may indicate a change in the communicative purpose. That is, in Example 6, we see an appearance of the phrase “post-traumatic stress syndrome”, not to define the concept, as did the beginning of Example 5 with “fictive indirectness”, but rather to apply the construct of “post-traumatic stress syndrome” to a series of plot events. Despite similar syntactic structures in phrases “post-traumatic stress syndrome” and “fictive indirectness” and their status as theoretical constructs, it appears that the Dimension 1 features around them may vary depending on the context in which they appear.

Biber et al., (1999), in their seminal grammatical study, suggest that abstract sentence subjects predominate in scientific writing. An abstract sentence subject in Example 5 is “interpreting fictive indirectness”, while one in Example 6 is “Caruth’s analysis of post-traumatic

stress syndrome". Both of these nouns are, arguably, epistemic sentence subjects, which MacDonald (1994) defines as ideological positions, past research, or abstractions used in the building of an argument. Epistemic sentence subjects, although by no means absent from literary criticism, have been found to occur more often in psychology writing (MacDonald, 1994; Wilder, 2005; Lake & Cortes, 2020). It is possible that looking for passages rich in negative-loading Dimension 1 features could aid in determining where epistemic sentence subjects occur in LCAs.

Example 7, taken from the highest-scoring Dimension 1 article, contains an application of theory to circumstances in the novel without defining any theoretical construct. Table 4.7 subdivides the dimensional features found in Example 7.

- (7) Freud (1949:18) **assumes** that people **have** a wish, usually, of course, unconscious, to die. This **indicates** that life **can** be a painful and exhausting process and that death **promises** release from this struggle. Zwelinzima **desires** peace, escape from stimulation. He cannot **imagine** any happiness without Thembeke next to him, **because** she **is** the one who actually helped him to overcome the feelings of childhood rejection and parental loss by being close to him all the time. He **is** overpowered by a realistic anxiety which **is** called fear. He **has** a fear of facing the world alone without his loved ones, especially his wife, whom he **thinks** he betrayed by allowing the hard-core amaMpondomise to pay lobola for a Bhaca girl to be his second wife. (en-282.txt)

Table 4.7. Dimension 1 features subdivided as % of word count in Example 7

Negative loading features:	Nouns	30	24.0%
	attributive adjectives	5	4.0%
	Prepositions	15	12.0%
	Sum		40.0%
Positive loading features:	The verb to be	5	4.0%
	Present tense verbs	7	5.6%
	Coordinating conjunction	3	2.4%
	That deletion	1	0.8%
	Can	2	1.6%
	Causal subordinating conjunctions	1	0.8%
	Sum		15.2%

Taking Examples 5-7 into account, it becomes possible to formulate a tentative hypothesis to explain the reasons for which Dimension 1 scores vary in LCAs. It appears that lower Dimension 1 scores may be associated with greater instances of elaborating upon theoretical constructs that require high levels of prepositions, attributive adjectives, and nouns. In the case of Example 5, this phenomenon takes place with such dense sentences as “[t]his technique is critical to Wendt’s approach to decolonizing **Pacific** histories, a methodology which recognizes a fluid boundary between history and fiction, thus taking up **imaginative narrative** as a **critical tool** through which to explore **psychological** and **experiential realism** (Wendt 1987).” Higher Dimension 1 scores may occur as a consequence of applying said concepts to summarized phenomena in the novels studied, as well as retelling and/or describing plot details. In Example 7, for a passage which applies a theoretical construct to a set of phenomena in a novel, positive loading features are almost twice as common as they are in the passage in Example 5, which defines a theoretical construct.

Concerning relations to other academic disciplines, Figure 4.2 shows that LCW scored at levels comparable to those reported for humanities prose by Biber (1988), while also higher than norms for natural science and medical academic prose. One detail that may separate academic medical science writing from LCW is that, in a study which plotted specific research article sections on the first five of Biber’s dimensions, Biber & Finegan (1994) found that Medical Science Methods sections scored lower on Dimension 1 than Introduction, Results, and Discussion sections, in part due to methods sections’ relative scarcity of present tense verbs. Conversely, as discussed earlier in section 4.1, none of the LCAs in English (or Spanish)

explicitly discussed analytical procedures. Thus, we find this to be one concrete detail that distinguishes LCW from prose in other academic disciplines.

4.2.3 Dimension 2: Narrative vs. Non-narrative Concerns

Dimension 2, described “as distinguishing between active, event-oriented discourse and more static, descriptive or expository types of discourse” (Biber, 1988, p. 109), is of particular relevance to this project, which aims to distinguish between discourse segments that retell and explain phenomena and events., a concept which we revisit in Section 6.1 of Chapter 6. The levels for LCW in relation to other registers are shown in Figure 4.3.

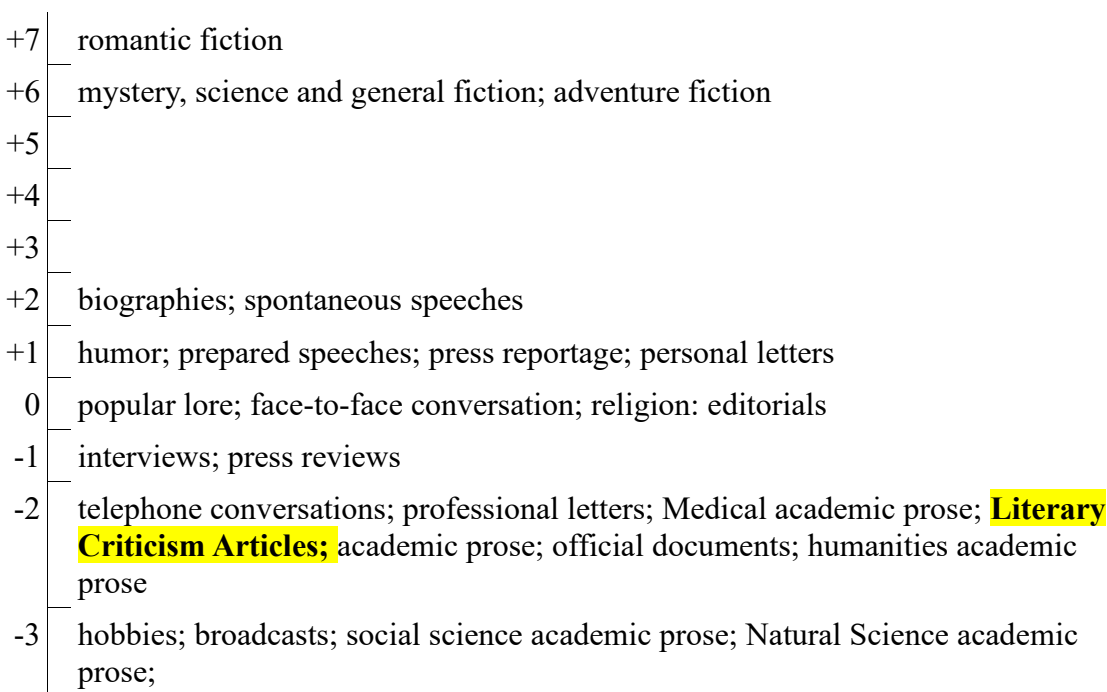


Figure 4.3. LCW articles in relation to other text varieties on Dimension 2 in English.

Mean Dimension 2 scores for English-language LCW score (-2.4) sat between the levels for professional letters (-2.2) and academic prose (-2.6). To explain how literary criticism articles would differ so markedly from works of literature regarding past tense usage, we note first that Biber et al. (1999), found that academic writing favors the usage of the present tense and the agentless passive voice. Indeed, fewer past tense verbs occurred in this corpus, compared to face-to-face conversations, and especially when compared to novels. A qualitative inspection of several articles' usage of Dimension 2 features suggests that a tendency may exist for what we might call, in this study, "Dimension 2 Deserts and Oases" in LCW. The term "desert" refers to a dearth of Dimension 2 features in a particular passage, while "oasis" refers to a passage rich in Dimension 2 features. To illustrate, the first paragraph in Example 9 comes from the introduction, while the second begins with a passage which establishes the historical context, a phase in literary criticism well attested in past English- (Fahnestock & Secor, 1991; Wilder, 2005; 2012) and Spanish-language (Viera Echevarría, 2014) studies.

- (9) Although the Matacão's mysterious appearance erases its links to heavy industry, to the rigs and pumps of oil production, this is not to **say** that, like plastic objects in everyday life, it is without historicity or geographic specificity. [...] This article therefore takes up Boetzkes and Pendakis's (2013) account of plastic as making $\langle \rangle$ (4), and seeks to elaborate upon **their** insight by reconnecting oil's production of material goods and ecological surplus with petroleum's infinite mutability and the industry's strategic fuelling of foreign debt. The first section examines Brazil's transformation from the 1960s onwards, moving from industrialization towards agri-exports as **funded** by petro-dollar debt mechanisms.

Brazil's history of neoliberalism

Ever since its colonization in 1500, Brazil's economic history has **been** structured by extractive policies and commodity manias, including the 17th-century sugar rush **followed** by a late 17th- and early 18th-century surge in gold mining; the late 19th-century rubber boom and subsequent emergence of 'Fordlandia' and other industrial rubber plantations in the early 20th century; and the late 19th- and 20th-century expansion of large-scale coffee plantations propping up Brazil's commodity export economy to this day. It **was** the period after the military coup of 1964, though, that **saw** the transformation of the Amazon rainforest into yet

another new commodity – and, indeed, waste (see Moore 2014, 38) – frontier. In the 1960s and 1970s the Brazilian military government actively **sought** out cheap international debt from the World Bank to initiate infrastructural and agribusiness projects in the underdeveloped north. **Rolled** out under the banner of ‘Operation Amazonia’, these projects **included** the development of new dams, cattle ranches, industries and mines – all of which **exacerbated** land clearing and created violent turf wars between indigenous settlers and predatory land speculators. The World Bank’s cheap loans **were** funded by petro-dollars: with the rising price of oil, OPEC (Organization of Petroleum Exporting Countries)’s producing nations **deposited** dollar surpluses into Wall Street banks that **developed** a series of financial mechanisms, such as Structural Adjustment Programmes, to recycle the money as debt. This, in turn, **financed** infrastructure projects throughout developing nations, and **led** to a boom in short-term but lucrative agricultural projects and export crops (see Niblett 2012, 26; Ortiz 2016, 611). With agro-industrialization **dominated** by the imperatives of ‘finance capital and its valuation of labor and nature in terms of short-term profitability’ (Ortiz 2016, 617), Brazil **was encouraged** to transition from an industrialized economy towards one based on high-profit agro-exports in order to service **its** debt repayments. (en-558.txt).

In the introduction phase, we have one token of a public verb (say), a third-person pronoun “their”, and a postnominal modifier “funded”. These together account for fewer than 1% of the text in a 327-word paragraph, indicating that the paragraph may be seen as a “desert”. The second paragraph in Example 9, however, contains 5% Dimension 2 features, more than a five-fold increase from the levels found in the previous paragraph, resembling more closely an “oasis”. Example 10 shows this tendency at play in two near paragraphs in another literary criticism article. The first in the set has two third-person pronouns and a past tense verb, whereas the second soars at 7.2% of its words logging as Dimension 2 features.

- (10) [...] As Richard Lane (2006, 37) usefully summarizes, Achebe’s achievement in *Things Fall Apart* lies in **his** resistance to Eurocentric “modes of analysis and representation”, especially the individualism characteristic of its literature, preferring instead to represent Umuofia from the perspective of a local metaphysics of communal identity, cyclical temporality and tribal spatiality. **He** thus produces, in John Thieme’s terms, a “revisionist fictional history” or “alternative historiography” (2001, 18, 19). [...] **Given** its comparable ambition and scope, *Time of White Horses* must be contextualized as a Palestinian

extension of such revisionist, or postcolonial, historical fiction.

[1 paragraph omitted]

Nasrallah's idealization of village life is indicated from the first sentence of *Time of White Horses*: <> (2012a, 3). While directly referring to Hamama, a white horse, a "mass of light" symbolizing the "free spirit", which is suddenly manifested in Hadiya, this sentence is also suggestive of Hadiya or Palestine itself (4). Hamama **had been** stolen two days earlier and, upon **her** arrival in **their** village, Hajj Mahmud and **his** son Khaled **set** out to retrieve **her** and restore the balance of nature that **had been** disrupted by this theft. <>, Khaled implores, <> (4). Shortly thereafter, Hamama's rightful owners arrive from a neighbouring village, and after an elaborate ritual of welcoming and exchange **she** is handed back to **them** as per custom. Both allegorizing and exemplifying the dynamic on which the rest of the narrative hinges, this opening episode posits a law of ownership sanctified by an organic, symbiotic relationship with nature or land, **policed** through the observance of custom and **threatened** only by abstract external agents who would claim possession by force. (en-556.txt)

Whereas most of the verbs in Paragraph 2 of Example 9 were in the past tense, the second paragraph in Example 10 differs. Rather than a contextual description of real-world history narrated primarily in the past tense, the second paragraph in Example 9 shifts between several verb tenses to describe events taken for granted in the fictional universe of the novel studied. To appreciate properly the extent to which this phenomenon occurs in Example 10, a tense-tracking diagram (Svalberg, 1995), which charts verbs' temporality as before, during, or after the narrator's point of view, was created. Svalberg (1995) defines story time as the point at which events take place in the story, whereas speaker time refers to the temporality perceived as current by the speaker. Figure 4.4. shows an arrangement of the verbs in Example 10 plotted in temporal relation to one another. Figure 4.4 demonstrates that the present tense occurs in both

speaker time and story time. However, when occurring in Speaker time, the present tense works to articulate the interpretations of the researcher.

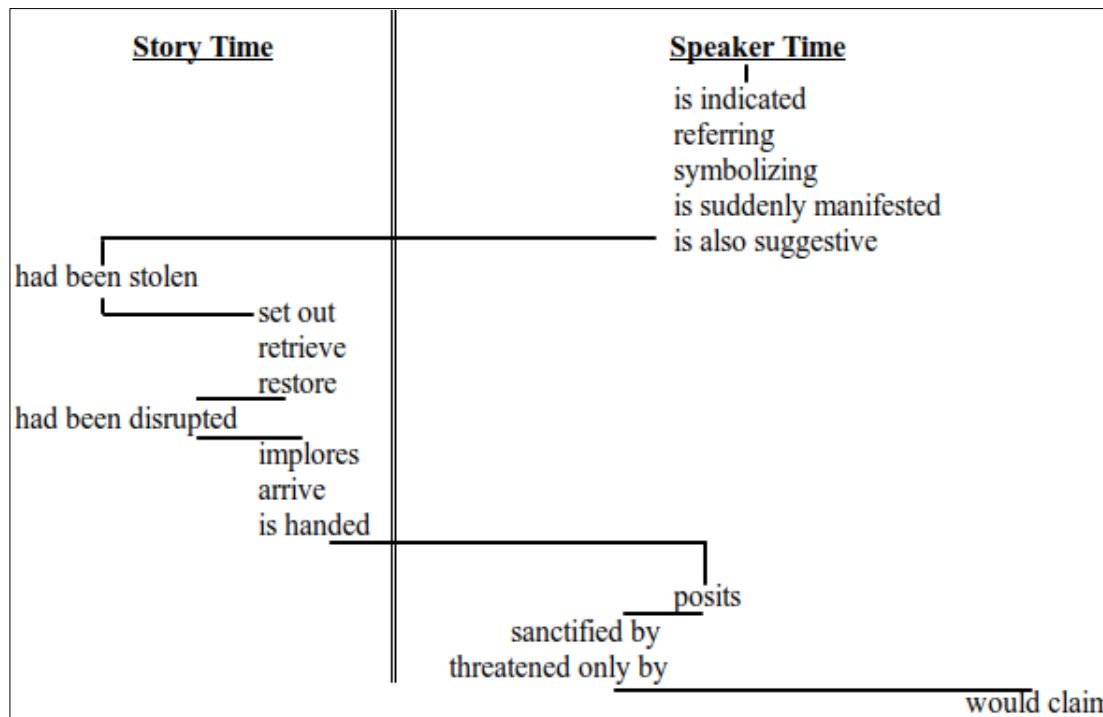


Figure 4.4. Tense tracking diagram of Example 10.

However, when used in story time, it refers to actions taken by the characters. The chart also reveals that variations on the past tense can also occur in Story and Speaker time.

A variation of the “Dimension 2 Oases” oasis appears to occur in LCW with paragraphs high in third-person pronouns and public verbs. In Example 11, we see the use of the third-person pronoun “He” referring to the author of the novel studied, in proximity to the public verbs “to argue” and “to suggest”.

- (11) [...] By tracing in minute detail the effects of Zionist colonization from the 1880s on traditional Palestinian village life, Nasrallah [**has provided**] in this novel the most expansive literary representation of the Nakba and its colonial prehistory we currently possess. **He** thus seeks to resolve what might be considered, in Franco

Moretti's (1994, 233) terms, the "symbolic problem" of the invisibility of the Nakba in global culture. To do so, **he** appropriates the European genre of historical fiction and re-inflects it with the Palestinian voice. I **argue** that this attempt results only in a contradictory, fragmented, riven form. [...] Although Nasrallah grounds his narrative of the destruction of Palestinian society during this period in oral history, such scale, in turn, necessitates the allegorical, mythic register **he** employs throughout. On the other hand, as if to offset the corresponding lack of interiority, **he** interjects a series of actual oral testimonies from survivors across the narrative via a compensatory historical register. [...] **Suggesting** the simultaneous necessity and impossibility of historical narrative, I conclude that *Time of White Horses* foregrounds the impossibility of aesthetic resolution in the Palestinian case without a national resolution in the form of return. (total paragraph length: 319 words (en-556.txt))

Example 12, taken from the highest scoring Dimension 2 article, is also a "Dimension 2 Oasis".

The pronoun "they" refers to the characters, and the "she" to the author of the novel. The token of "suggest" forms a hedging device around a component of the thesis statement occurring in the abstract. These third-person pronouns and public verbs are contributors to higher Dimension 2 scores.

- (12) Abstract: This article offers a feminist literary analysis of the gendered embodiment of shame in *Pompidou posse* by Sarah Lotz. [...] A close reading of the text reveals that the characters are exposed to overwhelming social messages of female culpability in a larger context that is rife with misogyny. As a result, **they** anticipate blame to such an extent that **they** blame **themselves** and internalise this blame as shame. By focusing on the bodies of the survivors, Lotz demonstrates the embodiment of shame, but **she** also **suggests** a corporeal challenge to silencing. The bodies of these characters speak loudly, albeit sometimes in the halting language of trauma, and **they** function to alert **them** to danger, to help **them** excavate memories that are made inaccessible and to **testify** to traumatic sexual assault. (en-520.txt)

The interaction between pronouns and public verbs on Dimension 2 may at times constitute integrated citation strategies (cf. Hyland, 1999). Hyland (1999), from a sample of qualitative social science and philosophy articles, generated a list of verbs used for negative, neutral, and positive integration of citations which used a scholar as the sentence subject (cf. MacDonald,

1994). Of the public verbs which Hyland (1999) identified as common triggers for integrated citations, “to argue” and “to suggest” occur 749 and 846 times, respectively. Further, of the 63 public verbs types occurring in the present corpus, the top two verbs are “to suggest” (20.6%) and “to argue” (18.3%), “say” (10%) and “explain” (9.7%), account for 58.7% of the public verb tokens. Thus, to a certain extent, Dimension 2 scores may serve as one criterion by which to detect integrated citations.

Concerning the relationship between LCW and other academic disciplines, Biber (1988) writes of humanities texts on Dimension 2 that they exhibit “a topical concern for concrete events and participants” (p. 193). The maximum score for humanities texts in the LOB corpus was approximately +5, but the maximum for LCW in this study was 1.3. These results suggest that the use of past tense verbs in LCW may be less frequent than in even a general pool of humanities writing. Indeed, unlike with Dimension 1 scores, Dimension 2 scores for LCW were on par with those of medical academic prose. Studying a separate collection of medical articles, Biber & Finegan (1994) found that Medical research article introductions scored significantly higher (-2.00) on Dimension 2 compared to Results, Methods, and Discussion sections (<-3.00). That a set of scientific article introductions could score more closely to LCW on Dimension 2 would explain why a schematic based in part on the Swalesian Create a Research Space (CARS) Model enjoyed a modest degree of success when applied to a collection of literary criticism abstracts (Tankó, 2017).

In summary, it appears that Dimension 2, when applied to LCW may measure both summaries of events in the past tense and the restating of integrated citations, although not

necessarily simultaneously. If a future MDA were carried out specific to LCW, it is possible that public verbs, third person pronouns, and past tense verbs might load on distinct dimensions.

4.2.4 Dimension 3: *Explicit vs. Situation-dependent reference*

Concerning elaboration and explicitness, this dimension is the only one in which literary criticism scores higher than all other texts in the three corpora analyzed by Biber (1988). Figure 4.5 shows the results.

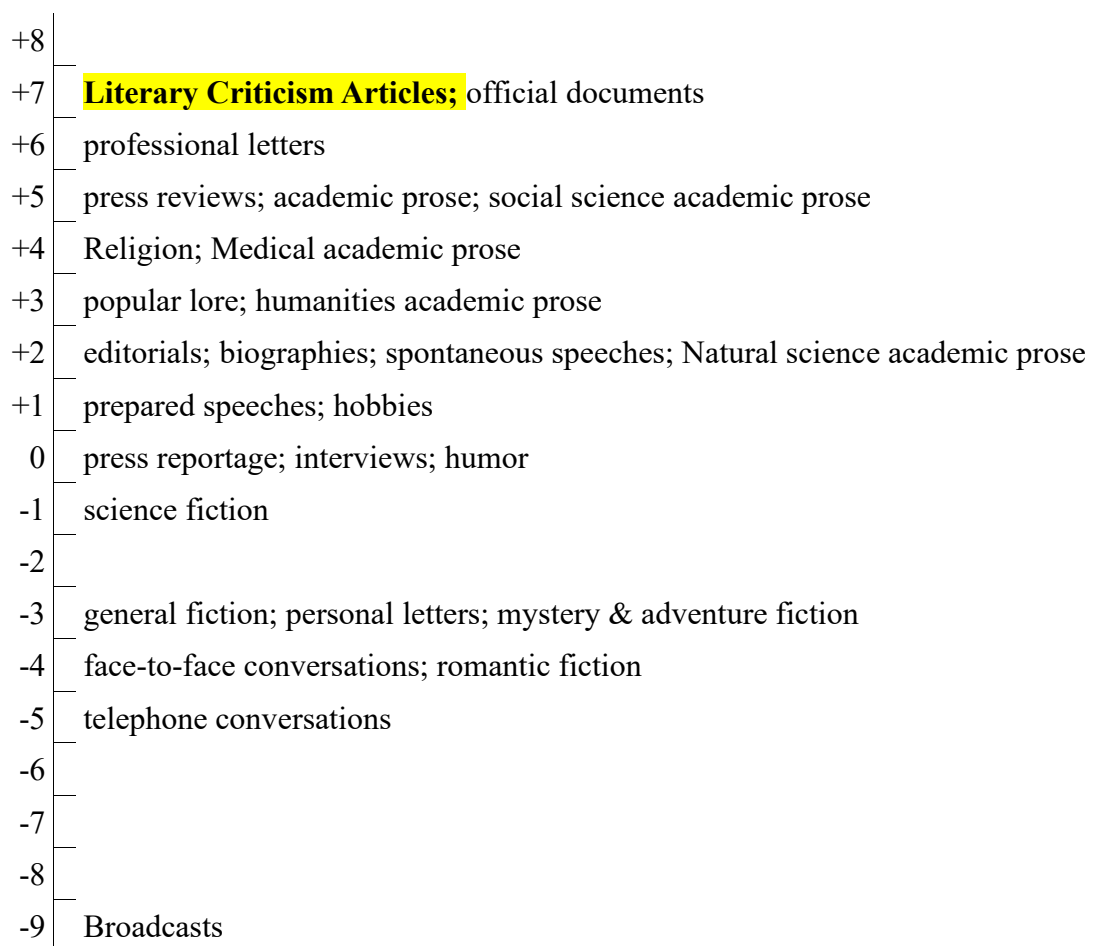


Figure 4.5. LCW articles in relation to other text varieties on Dimension 3 in English.

LCW appears to exceed the mean scores reported for other registers by means of precise, nominalization-based terminology and wh-clauses which overpower the generality invited by time and place adverbials. The fact that none of the examples had articles in the corpus that had Dimension 3 scores below 0 offers support for past remarks that LCW exhibits high levels of elaboration and explicitness (Biber & Gray, 2015). Biber (1988) remarks that Dimension 3 is useful at distinguishing between texts that are endophoric (exhibiting self-contained bits of information elaborated within the text) vs. exophoric (ones that require higher degrees of contextual inference to interpret) (cf., Halliday & Hassan, 1976). The positive loading features, which are phrasal coordination, nominalizations, wh- relative clauses, and pied piping constructions, add to a text's degree of explicit reference. The negative loading features, that is, adverbs, as well as time and place adverbials, appear to make the text more exophoric – or relate to reference that is more dependent on situations rather than explicit.

Example 13 comes from a low-scoring Dimension 3 passage, on par with dimension scores reported for academic prose (Biber, 1988). Its nominalizations are contrasted with adverbs and place adverbials (highlighted in yellow).

- (13) Let's start **at** the inquest. As many will recall, Nick's casual **reference** to legal proceedings— (137)—whipsaws Gatsby's plot in a **direction** first-time **readers** are not likely to expect. The plot **suddenly** becomes more Greek-tragic, and **yet also** sordid and half-unseen, a tale of garages, coffee joints, and ash heaps. But **even** my saying that we should < **here** demonstrates the complicated role newsmaking has played in the belated sequencing of Gatsby's own **narration**. **Here** Nick **abruptly** signals, **again** much in the manner of a glimpse, that **what** we have been reading, and **indeed** are about to be, is a counternarrative to a **never** fully realized public scandal that had run parallel in the tabloids and the courts. And **soon** he follows with another jarring note, another quick shift into the past tense. < jumps **out** (138), much as Myrtle Wilson herself is about to have jumped **out**, and melodrama starts its shadowing effect **even** before we arrive at the novel's finale. (en-377.txt)

Indeed, the passage's negative-loading features outnumber positive-loading ones as a consequence of shifting temporal referents frequently while narrating plot details in the historical present.

In Example 14, one sees a more measured degree of conceptual relatedness within a single paragraph. The literary critic embarks on a deep elaboration of plot details which branches out into societal implications.

- (14) If Paris became an African diasporic literary capital in the 1950s and a communal site of exile **particularly** for African American **writers**, it **also** became an important alternative for jazz musicians who sought **acceptance** as artists and **opportunities** as professional **entertainers**. Jazz became **more** popular in France after World War II partly because it had become a “cultural symbol of **antifascism**” during the **Occupation** (Stovall 134). African American exiles **also** became **politically** appealing to French intellectuals, as their **resistance** to US **racism** corresponded with growing **skepticism** on the French Left toward US Cold War foreign **policy**. Jazz symbolized **what** was most despicable about the US as **well** as [**what**] was most appealing since blacks were viewed as both “victims of American **brutality** and heralds of the New World’s choicest offerings” (136). (en.475.txt)

Using a balance of negative- and positive-loading Dimension 3 features, the literary critic appears to establish a manifestation of the Context *topos* (cf. Fahnestock & Secor, 1991), defined as a recurring rhetorical strategy in literary criticism writing that supplies real-world contextual information relevant to an understanding of the novel studied. The article does so through a combination of explicitness and situation-dependent language to provide information about the significance of the Paris arts scene in the promotion of jazz as part of a larger argument about the novel.

Example 15, from a high-scoring Dimension 3 passage, puts forth even more explicit reference to abstract phenomena.

- (15) The **introduction** is meant to draw, redirect and refocus people's **attention** at the terminus. It is meant to make sure that the assumed **offender** and other people do not miss **out** this verbal avalanche and other invectives. It is **also** an **introduction** that is loaded with anger and **emotion** as **well** as discourses of **intimidation**. The poem is **strategically** placed **at** the beginning of the novel to set the stage for the novel's uncompromising **stance**. In the novel, it is **indeed** intended to be a vituperative and vitiating attack on systems that dehumanise people. The reasons that plunge most of the characters like Kundai, Magi, Maud, Vincent and others into an unfathomable lapse of **vitality** and **immorality** as **well** as **degeneration** into ideological **captivity** and mass neurosis are captured in this introductory poem. Kundai and Magi are two young women in the novel **who represent** the **majority** of the University of Zimbabwe's female **students**. As a result of reduced and **yet** much-needed **government** support, in the form of grants and loans, they resort to **prostitution** in order to survive. Vincent is an unemployed man, who survives by selling outlawed drugs like dagga while Maud is a widow who earns a living through cross-border **business**. (en-277.txt)

The first three sentences all refer to the introduction of a novel as they elaborate on what the researcher argues is the introduction's greater purpose. It bears mentioning that nominalizations are the most highly represented Dimension 3 feature. Table 4.8 shows the average Z score for normalized frequency counts of each Dimension 3 feature for each text file in the corpus.

Table 4.8. Z-scores for normalized frequency counts of Dimension 3 features in English corpus

Group	Feature	Mean	SD
Positive Features	Nominalizations	4.3	1.0
	Pied piper clauses	0.9	1.0
	WH relative clauses on subject positions	0.5	0.7
	phrasal coordination	0.4	0.6
	WH relative clauses on object positions	0.3	0.4
Table 4.8 (Continued)			
Negative Features	Place Adverbials	0.2	0.5
	Time Adverbials	-0.7	0.4
	Adverbs	-0.8	0.4

Nominalizations, as a subset of nouns, contribute to the density of nouns in a given sentence -- an attested Dimension 1 feature. Just as Section 4.2.2 of Chapter 4 suggested that high Dimension 1 scores may be indicative of retelling plot events while low ones may lend themselves to explaining scholarly phenomena, low Dimension 3 scores may signal the presence of retelling while high ones may also be associated with explaining abstract entities.

In relation to other disciplines, LCW's scores are higher than all other types of academic prose studied by Biber (1988), but comparable to scores reported for Medical Discussion and Introduction sections (Biber & Finegan, 1994). That both Discussion and Introduction sections would resemble LCW to a slightly greater degree than other sections may occur as a consequence of Discussion and Introduction sections' having been found to share communicative purposes that involve descriptions of and generalizations about extant knowledge on a given topic (Kanoksilapatham, 2005).

4.2.5 Dimension 4: Overt Expression of Persuasion

Biber (1988) writes of Dimension 4 that it represents the extent to which "persuasion is marked overtly, whether overt marking of the speaker's own point of view, or an assessment of the advisability or likelihood of an event presented to persuade the addressee" (p. 111). Figure 4.6 shows how LCW scored in relation to other text varieties. With the exception of broadcasts, LCW scored lower on this dimension than all other registers examined by Biber (1988). As a prior criterion by which to detect where one might find any semblance of overt persuasion in an LCA, we consider that the final phases of some Spanish literary criticism presentations were found to include at least some level of persuasion (Viera Echevarría, 2014).

+3	professional letters; editorials
+2	romantic fiction; hobbies; personal letters
+1	interviews; general fiction; telephone conversations; prepared speeches; spontaneous speeches; religion
0	official documents; face-to-face conversations; humor; popular lore; academic prose; biographies; mystery and science fiction; press reportage; humanities academic prose
-1	adventure fiction
-2	press reviews; medical academic prose; social science academic prose
-3	Natural science academic prose
-4	Literary Criticism Articles; broadcasts

Figure 4.6.. LCW articles in relation to other text varieties on Dimension 4 in English.

With this knowledge in mind, Example 16 was selected from the conclusion of a low-scoring Dimension 4 document.

- (16) [...] In an attempt **to halt** the epidemiological wave, a powerful insecticide is sprayed that only serves **to kill** all the Amazon's birds, in a scene redolent of Rachel Carson's 1962 account of the effect of DDT on birds in *Silent Spring* (Heise 2008, 103). [...] [...] This insistence on a brief image of nuclear family bliss strikes an inauthentic note of clichéd individualized happiness, a peculiarly regressive mode of sociality and bundle of relations that are set against the collective re-imagining required **to resist** neoliberal development. [...] (en-558.txt).

The only Dimension 4 features occurring in Example 16 are three infinitives. Biber (1988) found infinitives to associate with the expression of persuasion due to their co-occurrence with modals of obligation. In Example 17, a medium-scoring Dimension 4 conclusion, we see the infinitive “to remove” and the conditional subordinator “if”.

- (17) Still, Chesnutt has his historical facts wrong in the case of the real Dr. Millers. In the actual Wilmington riot, whites specifically aimed **to remove** economically potent, property-owning African Americans. Josh Green has more insight into the

logic of Southern racism, then, than the determinedly optimistic Miller, or Chesnutt for that matter, **if** we read Miller's exemption from mob violence as an instance of Chesnutt's own fantasy of racial conciliation defined in superlative economic terms. [...] (en-376.txt)

Indeed, the latter feature appears to foreground a perspective which the author sees as correct vis-a-vis a less correct one by a previous, more "optimistic" scholar. However, no features characteristic of persuasion within an MDA framework appear for the remainder of the paragraph.

Example 18 contains a conclusion from a high-scoring Dimension 4 article.

- (18) This Africanist approach **should** explore and/or create possibilities of enriching the intellectual terrain of literary study by creating a healthy climate for robust, but positive academic debate among African literary scholars. African scholars cannot, therefore, remain unaffected by issues of socio-cultural concern in our era, and as a result literary study **should** also take note of the issues of culture and identity since these have become worldwide trends. To Africans, however, all charters and policies **will** be meaningful when their languages and cultures also become mechanisms **to relocate** African culture in the centre of national, continental and international developments and scholarships. (en-399.txt)

The modals "should" and "will" are followed by an infinitive to produce a more overt attempt to persuade readers of the author's vision for scholarly conversations on literary productions from and pertaining to the African continent.

Argumentation strategies were found to occur in the Introduction, Body, and Conclusion phases of several Spanish literary criticism conference presentations (Viera Echevarría, 2014). As such, the fact that so few overtly persuasive features occurred in LCW is curious. LCW exhibited even fewer persuasive features than all other sciences in Biber (1988), but more than medical results sections (Biber & Finegan, 1994). As Biber (1988) contends, "studies that are more experimental or empirical in nature (natural sciences, social sciences, etc.) depend less on the logical comparison of alternatives, and the use of persuasive form" (p. 194). A possible

explanation for the discrepancy between Biber's remarks on humanities writing and those of Viera Echevarría is considered, in Section 4.2.7, where we explore the possibility that adjective and verb complement clauses may act as "covert" expressions of persuasion. First, however, we will discuss Dimension 5.

4.2.6 Dimension 5: Abstract vs. Non-abstract information

Of Dimension 5, Biber (1988) writes that:

"these forms are all used to present propositions with reduced emphasis on the agent, either demoting the agent to object position or eliding the agent altogether. They are used to give prominence to the patient of the verb, the entity acted upon, which is typically a non-animate referent and is often an abstract concept rather than a concrete referent. Passives are frequently used in procedural discourse, where the same agent is presupposed across several clauses and the specific agent of a clause is not important to the discourse purpose" (p. 112).

Figure 4.7 shows that LCW scores higher on this dimension than most other texts, with the exceptions of hard sciences, academic prose, and official documents.

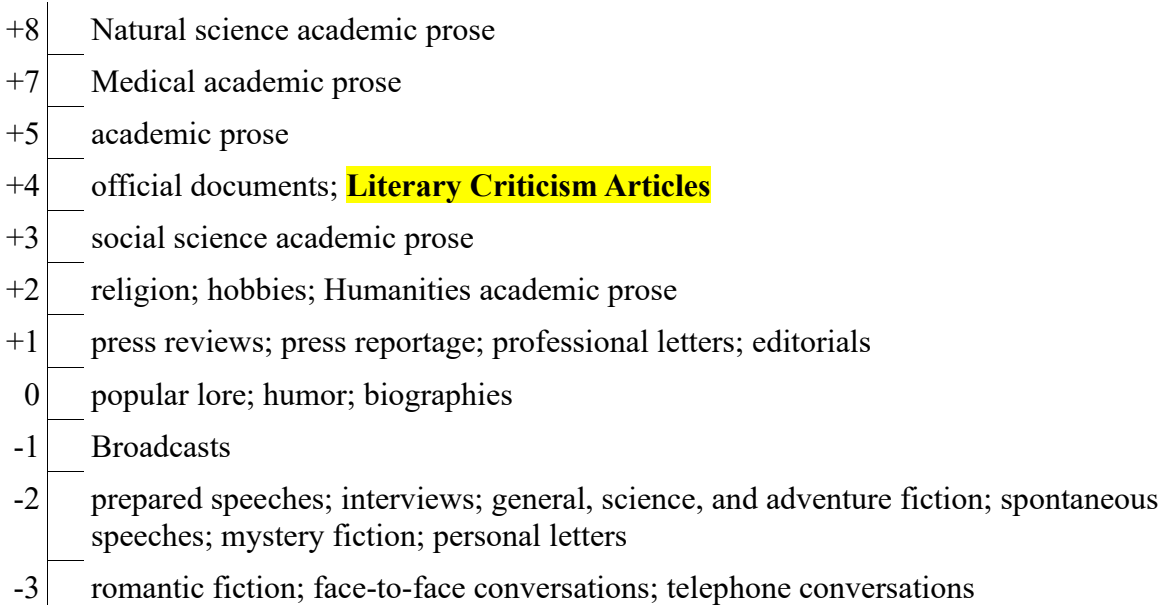


Figure 4.7. LCW articles in relation to other text varieties on Dimension 5 in English

Indeed, one finds a great deal of emphasis on the patient of the verb (mean normalized frequency of agentless passives 7.9, SD = 2.8) with less mention of the agents (overt passive =2, SD= 0.9). The lowest scoring document, from which an excerpt of a paragraph appears in Example 19, contains just one agentless passive.

- (19) In other cases, such as when the characters **are disconnected** from the rest of the family and their culture, they have to find their healing in what rituals they can. (en-466.txt)

In line with Biber's (1988) observations on the ubiquity of abstract patients in passive voice constructions, Example 20's patients in passive voice constructions include "Nobantu's loss of touch with reality", "this wish", "these wishes", along with the nonabstract NP "a Bhaca girl".

- (20) Nobantu's loss of touch with reality can **be interpreted** as being a punishment from the ancestors of the amaMpondomise, as she never respected them after marrying Zwelinzima. In Nobantu's development there is an association between

wishes and fears. After getting involved with Zwelinzima, she had a wish of getting married to him, but this wish **was destroyed** when she heard the news that Zwelinzima was a chief. Her fear was that she was a daughter of a commoner. She was not from the royal family, so her wish might prove futile because of her lover's status. These wishes and fears that are part of a specific time period in the person's life development **are carried over** into the next stage, adulthood. Nobantu's wishes and fears drove her to behave strangely in the Mpondomise homestead. She did not want to lose Zwelinzima by sharing him with another wife, a Bhaca girl. At the same time she could see that the traditionalist would never be happy **until** a Bhaca girl **was brought** into the home. The conflict between Zwelinzima and the amaMpondomise and between the amaMpondomise and Nobantu intensified after she killed the snake. This leads to Nobantu's neurotic symptoms and psychotic withdrawal from reality. The above-mentioned symptoms express Nobantu's unconscious internal conflict between self-assertion and conformation. (en-282.txt)

Abstract verb subjects have been found to predominate in academic prose (Biber et al., 1999).

Example 20 is no exception when one uses the tokens of the agentless passive as a focal point.

Additionally, the importance of a variant on the passive construction "can be interpreted as" as one used to express stance will be discussed in Chapters 5 and 6. For now, we suggest that it could function as a discipline-specific, covert means of expressing stance not measured by Dimension 4 scores.

Example 21 comes from a high-scoring Dimension 5 passage.

- (21) Though the essay "The White Negro" (1957) is often considered Mailer's core statement on existentialism, it wasn't **until** *An American Dream* (1965) that Mailer's distinctly American existential vision came to full fruition. *An American Dream* sees the manifesto-like dogmatism of "The White Negro" and several essays from *The Presidential Papers*—all of which champion violent masculine agency as a form of socio-political protest—transposed to the more subtle literary devices of fiction. And while past feminist critics, such as Mary Ellmann, Kate Millett, and Judith Fetterley, have correctly taken Mailer to task for his more overt representations of misogyny in *An American Dream*, there have been few efforts to reveal the moments in the novel that complicate such readings, those instances that introduce a greater complexity to Mailer's gender politics in the novel and to his existential vision as a whole. **However**, in *Sexual Politics* (1970), Millett acknowledged that, though Mailer was, indeed, a "prisoner of the virility cult," he was also <> (314). Millett subsumes Mailer's sexually charged existentialism

under the clever portmanteau “sexistentialism,” which she defines as a $\langle \rangle$ in terms of $\langle \rangle$ (327). The term “sexistentialism” will be used, in the present article, to refer to the virile, misogynistic existential project (“[****]ing as conquest”) **undertaken by** Mailer’s protagonist in the first half of *An American Dream*. **However**, I will demonstrate how the sexistentialism of the first half of the novel **is later destabilized by** the broader scope of Mailer’s existential vision in the second, which sees the dissolution of the protagonist’s sexistential project in the face of the absurd, emasculating power of post-war corporate America. **Thus**, rather than reading *An American Dream* as an inherently misogynistic text, I argue that Mailer uses representations of misogyny to draw critical parallels between two ostensibly separate—though in actuality intimately connected—exploitative practices: Mailer depicts Rojack’s slow realization that his own sexistential project of $\langle \rangle$ **is motivated by** the same exploitative practices that sustain the expansionistic capitalism that underlies the American Dream. (en-470.txt)

Example 21 stands out for its passive constructions with marked agents, e.g., “Mailer’s protagonist”, “the broader scope of Mailer’s existential vision”, etc., in addition to the conjuncts “until”, “However”, and “Thus”.

In all, it appears that the some of the Dimension 5 features in Examples 19-21 are used to make abstract inferences. A lapse therein, conversely, may indicate that the author is summarizing or narrating past information. This tendency may be a legacy feature of philology treatises on the Biblical corpus. Turner (2015) argues that the discourse practices in these treatises laid the groundwork for the modern literary criticism paper. Example 22 contains a translation of a paragraph by early philologist Origen of Alexandria. The second sentence contains a clausal-level summary of events written in the Gospel of Matthew. The paragraph argues that some passages in the Gospels are more suited toward figurative, rather than literal, interpretations.

- (22) And what need is there to say more, since those who are not altogether blind can collect countless instances of a similar kind recorded as having occurred, but which did not literally take place? Nay, the Gospels themselves are **filled** with the same kind of narratives; **e.g.**, the devil leading Jesus up into a high mountain, in order to show him from thence the kingdoms of the whole world, and the glory of

them. For who is there among those who do not read such accounts carelessly, that would not condemn those who think that with the eye of the body--which requires a lofty height in order that the parts lying (immediately) under and adjacent **may be seen**--the kingdoms of the Persians, and Scythians, and Indians, and Parthians, **were beheld**, and the manner in which their princes **are glorified** among men? And the attentive reader may notice in the Gospels innumerable other passages like these, so that he will **be convinced** that in the histories that **are literally recorded**, circumstances that did not occur **are inserted**. [De principiis. 4.3.1.]. (cited in Grant & Tracy, 1984).

A retelling of events narrated in the Gospel of Matthew begins with the conjunct “e.g”. Much like the techniques observed in Examples 19-21, sentences with the agentless passives appear to take place at moments where more abstract reasoning manifests. Of course, Example 22 also holds other deprecated features of textual criticism writing, such as mono-gender, rather than dual-gender, reference, in the phrases “glorified among men” and “he will be convinced”. Taking these details into account, it appears that passive voice strategies in tandem with conjuncts for abstract reasoning may be a rhetorical strategy dating back hundreds of years in the genre.

In relation to other disciplines, LCW scored higher on Dimension 5 than social science and even humanities academic prose, but lower than natural science and medical academic prose. Regarding disciplinary variation on Dimension 5, Biber (1988) argues that

“the extremely abstract form for scientific and engineering prose might also reflect the linguistic norms that are explicitly taught to scientists and engineers: that empirical studies are factual, and therefor faceless and agentless. Conversely, humanists are taught and teach that passives are dispreferred constructions and that good writing is active. In fact, all academic prose subgenres are quite abstract in form when compared to the full range of English genres, but technical and scientific prose represents extreme use of these forms” (p. 194).

Of interest is that LCW scored lower than humanities writing on Dimension 4, but higher than humanities on Dimension 5, indeed, favoring more passive voice construction. Moreover, while LCW Dimension 5 Scores sit close to those reported for Medical research article introductions, Dimension 5 scores for medical Methods sections (Biber & Finegan, 1994) were noticeably denser with passive voice constructions. A possible explanation for this phenomenon lies in Section 5.4.2, wherein we discuss the formulaic phrases “can be read as a” and “can be seen as” and their role in expressing the viewpoint of the author.

4.2.7 Dimension 6: On-line elaboration of information

Of Dimension 6, Biber (1988) proposes that “*that* (italics mine) complements to verbs and adjectives can be used for elaboration of information relative to the personal stance of the speaker, introducing an affective component into this dimension” (p. 114). Biber argues that information is “elaborated” via verb- (“suggests that X”) and adjective complement (“it is possible that”) clauses, as well as relative clauses. The fixtagging rendered Dimension 6 scores for LCAs further down to the ranks of two other registers before arriving at levels comparable with religious texts, as seen in Figure 4.8.

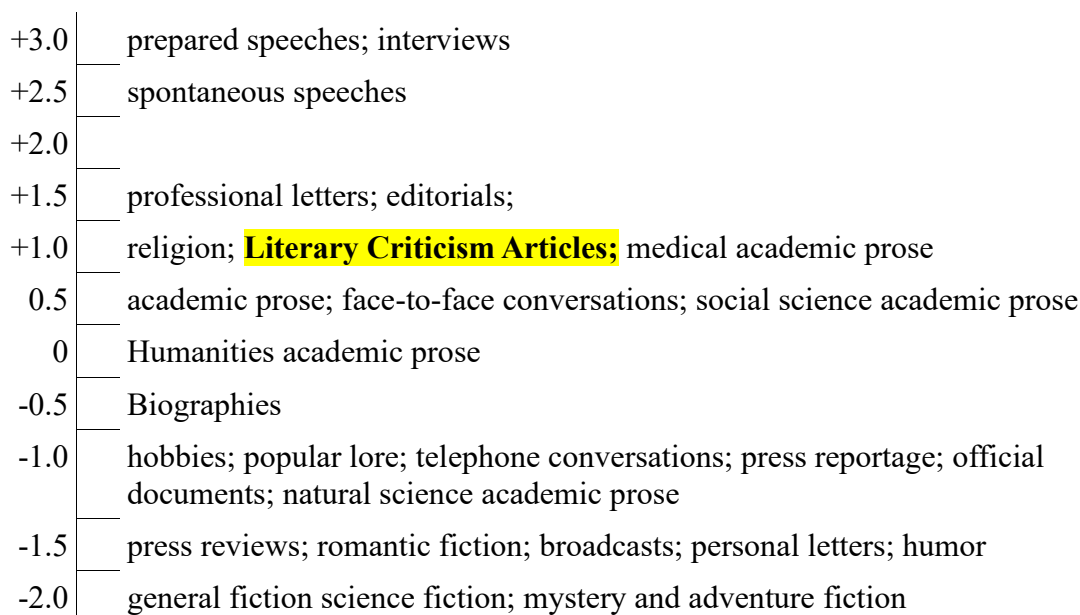


Figure 4.8. LCW articles in relation to other text varieties on Dimension 6 in English.

Indeed, it appears that Dimension 6 allows us to see the ways in which the literary critic can alternate between the points of view of her or his own, or those of the characters or other scholars.

Example 23 contains a passage from a low-scoring Dimension 6 article.

- (23) [...] Alienated from his work performing a routine task on an assembly line for a large electronics company, and living in the massive, dehumanizing housing complex owned by the same corporation, which plans to expand to Mars (137–8), Jack Bohlen initially improves when he escapes Earth’s conditions for Mars’s primitive simplicity, but his schizophrenia soon worsens when **these** conditions are recreated by Arnie Kott. Otto **laments that** \diamond and **that** \diamond on Mars as well (95). [...] (Total passage length: 158) (en-472.txt)

Example 23 involves two verb complement clauses which are connected to the character Otto’s “lament[ing]”. This verb portrays a fictional attitude as negative in the context of the larger argument that Philip K. Dick was a critic of consumerism and its concomitant generation of waste products.

Example 24, from a medium-scoring Dimension 6 passage, contains more Dimension 6 features.

- (24) [...] He interpolates a vast array of contrasting, oppositional elements of national cultures as his identity, but **this** interpolation is revealed in the text as absurd, impossible, and ugly—a psychological and physical deformity. [...] While Gilroy’s project sets out to reconstitute a working theory of universalism, he acknowledges **that** the contemporary political environment has all but occluded the value of theories of cultural integration. [...] **This** conversation is very much at play in Alfau’s work. Although written in the 1940s, Chromos expresses a pronounced sense of disenchantment akin to the \diamond and \diamond perceptions of multiculturalism **that** Gilroy describes. [...] Indeed, Chromos never posits a productive sense of universalism; in fact, it is the very act of negotiating an alternative to either a national or universal consciousness **that** Alfau struggles with throughout the text. (en-482.txt)

The demonstrative determiners in Example 24 co-occur with nominalizations, which constitute an important component of elaboration strategies in academic prose (Biber, 1988). Two relative clauses with object gaps in the form of “perceptions of multiculturalism that Gilroy describes” and “universal consciousness that Alfau struggles with” occur. Finally, one notices a verb complement clause (“he acknowledges that”) used to put forth the second of two facets of another researcher’s assertions.

Example 25 comes from a high Dimension 6-scoring passage.

- (25) [...] Even though Lucas is the grandson (and great-grandson) of Carothers McCaslin, the McCaslin family patriarch, he cannot claim **this** white identity and, as a result, is regarded by every other character in the novel as fundamentally black. **This** racial essentialism, which ignores the nuances of Lucas’s biracial background, conforms to a long-standing logic of racial identity in the United States that holds **that** a single, non-white ancestor, or “one drop” of non-white blood, is enough to categorize **that** person as black. Indeed, it is **this** principle more than any other that organizes the decidedly convoluted McCaslin family tree, determining who is a member of the family and who is the family’s property. [...] In **this** regard, **this** symbolic law about the circulation of different kinds of blood is analogous to an economic precept—Gresham’s Law—that explains the circulation of different kinds of money. Put simply, Gresham’s Law states **that** \diamond . [...] **This** anxiety over whether a particular kind of money should be spent or stored, points toward the second of Lucas’s crises, as he spends a remarkable

amount of time and energy trying to save—and enlarge—his secret cache of gold coins. In what follows, I explore the connection between **these** two laws of circulation in Faulkner’s text, situating the novel’s sustained interest in both biological and financial inheritance within a broader history of the relationship between slavery and the monetization of gold. I argue **that** there is a homology, in *Go Down, Moses*, between the operation of the one-drop rule and the principles identified by Gresham’s Law that brings to the fore Faulkner’s interest in the cultural politics of inheriting both family fortunes and racial identities. Lucas attempts to reify his nebulous racial identity by fetishizing (and hoarding) a particular form of money (gold coins, in **this** case) **that** he associates with his white, McCaslin ancestry, substituting the material properties of golden money for the immaterial whiteness that has such a strong significance for his personal identity. [...] (en-462.txt)

Five of the nine demonstrative determiner noun phrases refer to abstract ideological entities:

“this white identity”, “This racial essentialism”, “this principle”, “this symbolic law”, and “these two laws of circulation”. An object-gap relative clause also manifests as “a particular form of money that he associates”. Finally, a variation in stance occurs in the verb complement clauses. The “one-drop rule”, for instance, “holds that”. Gresham’s law “states that”. However, the author, when referring to his own perspective, writes “I argue that”.

In relation to other disciplines, LCW scored higher than all other genres of academic prose, including those reported for humanities initially. Verb complement clauses were the second-highest contributing Dimension 6 feature in the corpus (a close second to object-gap relative clauses). One factor at play may lie in the fact that the number of bibliography entries per research article in fields such as engineering (Ucar et al., 2014) and philosophy (Petrovich, 2018) has been shown to increase substantially within the past 40 years. As such, more references cited in LCAs, which like philosophy, are a humanities field, may necessitate more verb complement clauses per text. In support of this possibility, Table 4.9 shows the common verb phrases which initiate verb complement clauses, along with tokens.

Table 4.9. Top 10 verb phrases triggering verb complement clauses

Rank	Word	Frequency
1	argues	298
2	suggests	232
3	is	228
4	suggest	140
5	argue	139
6	notes	127
7	explains	87
8	writes	78
9	claims	71
10	suggesting	60

It is interesting to note that many of the verbs in Table 4.9 are public verbs, which were a positive-loading Dimension 2 feature. However, the interfactor correlation table revealed that there is no significant relationship between Dimension 2 scores and those of Dimension 6. These results suggest that Dimension 2 may communicate both past tense verbs and private verbs in the present to such an extent that make it equally likely to narrate plot events or summarize fictional-historical details. As such, the value of Dimension 6 as being able to uniquely capture complement clauses is highlighted.

4.3 Spanish MDA results

The present section addresses the second portion of Research Question #1 through the results and discussion of the MDA carried out in Spanish using Venegas' (2010) metrics for MDAs of textual Spanish-language corpora. Venegas' (2010) analytical approach uses the 5 highest loading linguistic features on the five dimensions devised by Parodi (2005) created to

analyze texts in Spanish. Due to Venegas' (2010) methodological transparency, the methods used in his study were chosen for the present dissertation. Section 4.3 contains results and Dimension scores for the Spanish-language corpus. Section 4.3.1-4.3.5 discuss the results of Dimensions 1-5 and their implications on the Spanish corpora. Lastly, Section 4.3.6 offers some concluding remarks.

For illustrative purposes, Table 4.10 shows the steps by which mean frequency counts for linguistic features are translated into dimensional density scores when multiplied by the dimensional loadings used by Parodi (2005) and Venegas (2010). Actual dimensional density scores (DDS) calculated based on the entire dataset rather than a single mean score appear in Table 4.11

Table 4.10. Mean and (SD) of linguistic feature frequency counts, loadings, and products of linguistic features

Dim.	Feature	Counts*	Loadings	(Counts × Loadings)*	Sum	× 5	Mean DDS
1	Cause/effect adverbial clauses	16.8 (9.1)	0.945	15.8 (8.6)	65.3	326.5	0.058
	Time adverbs	18.4 (10.5)	0.934	17.1 (9.8)			
	Negation adverbs	34.3 (17.4)	0.928	31.9 (16.1)			
	Second-person singular pronouns	0.1 (0.5)	0.911	0.1 (0.5)			
	First-person singular pronouns	0.4 (1.3)	0.823	0.4 (1.1)			
2	Second-person singular pronouns	0.1 (0.8)	0.842	0.1 (0.7)	35.8	179	0.032
	First-person singular pronouns	0.4 (1.3)	0.828	0.4 (1.1)			
	Periphrastic future	0.9 (1.5)	0.823	0.8 (1.2)			
	Imperfect past	19.1 (21.1)	0.820	15.6 (17.3)			
	Plural pronouns	26.6 (14.9)	0.708	18.9 (10.5)			
3	Private verbs	2.4 (2.2)	0.824	2 (1.8)	26.3	131.5	0.023
	First-person singular pronouns	2.4 (3.8)	0.789	1.9 (3)			
	Preterit	22.8 (22.7)	0.705	16.1 (16)			
	Modal verbs of volition	1.7 (2.1)	0.655	1.1 (1.4)			
	First person singular inflections	8.2 (8.8)	0.640	5.2 (5.6)			
4	Active forms of the verb ser (to be, durative)	29.6 (14.7)	0.671	19.8 (9.9)	64.1	320.5	0.057
	Hedging verbs	1.6 (2.3)	0.656	1.0 (1.5)			
	Modal verbs of possibility	15.3 (8.7)	0.641	9.8 (5.6)			
	adverbs of manner	33.3 (20.4)	0.606	20.2 (12.4)			
	Predicative adjectives	23.6 (12)	0.565	13.3 (6.8)			
5	Modal verbs of obligation	2.5 (2.6)	0.496	1.2 (1.3)	167.3	836.5	0.149
	Subjunctive mood	19.9 (11.7)	0.494	9.8 (5.8)			
	Nominalizations	165.7 (70.6)	0.456	75.6 (32.2)			
	Participles in adjectival functions	36.6 (18.4)	0.413	15.1 (7.6)			
	prepositional phrases as noun complements	158.9 (62.5)	0.413	65.6 (25.8)			

Table 4.11 contains the mean DDSs for the corpora analyzed in Venegas (2010), as well as the values for the corpus of literary criticism articles. These scores report how many of the linguistic features associated with each dimension score occurred in a given corpus. For example, the score of .175 on Dimension 2: Narrative Focus for Latin American Literature suggests that it contains high quantities of verbs conjugated in the past imperfect compared to corpora of other genres, e.g., .021 for public political speeches. Conversely, Latin American Literature's score of 0.79 on Dimension 5: Informational Focus indicates that the novels in the corpus contain prose that is less informationally dense compared to the Corpus PUCV-2006 (.172) which is comprised of university textbooks.

Table 4.11. Literary Criticism Articles in comparison with other registers from Venegas (2010)

Dimension	1	2	3	4	5
Corpus PUCV-2006	0.062	0.039	0.049	0.087	0.172
Oral Didactic	0.215	0.109	0.135	0.055	0.044
Latin-American Literature	0.163	0.175	0.258	0.068	0.079
Public Politics	0.050	0.021	0.023	0.060	0.175
Scientific Research Articles	0.014	0.015	0.032	0.032	0.149
LCW	0.059	0.031	0.024	0.056	0.150

Table 4.12 contains the interfactor correlations for the dimensional density scores in the Spanish LCAs. The data were not normally distributed. Thus, a nonparametric Spearman's Rho, rather than a Pearson's R, was calculated for the correlations between each factor. Calculating interfactor correlations is important because redundant variables in statistical analysis leave open a greater potential for misleading results. One strategy recognized as minimizing its effects is to inspect correlations between variables to ensure that no two in a given model have significant correlation coefficients with absolute values greater than 0.7 (Slinker & Glantz, 1985).

Table 4.12. Interfactor correlations for dimensional density scores in the Spanish LCAs

	Dimension 1	Dimension 2	Dimension 3	Dimension 4	Dimension 5
Dimension 1	1.00				
Dimension 2	0.22*	1.00			
Dimension 3	0.18*	0.42**	1,00		
Dimension 4	0.55**	n.s.	n.s.	1.00	
Dimension 5	-0.29**	-0.35**	n.s.	-0.24*	1,00

*p < 0.01

**p < 0.001

Table 4.12 reveals that no two factors were correlated at unacceptable levels. In the following sections, we will turn to discussions of how LCW's particularities are manifest through the analysis.

4.3.1 Dimension 1: Contextual and Interactive Focus

Within the parameters of the present analytical approach, Dimension 1 *Contextual and Interactive Focus*, used by Venegas (2010) has been described as measuring the extent to which a text contains a series of events, signaled by deictic reference to space, time, and demonstrativeness, communicated to a perceived reader or listener. That is, texts that score high on Dimension 1 tend to resemble spoken registers in that they contain a rich degree of discourse markers meant to clarify and qualify statements according to the receiver of the message (Parodi, 2005).

Figure 4.9 displays how Literary Criticism articles scored in relation to other genres studied by Venegas (2010). Note that the Pontifical Catholic University of Valparaíso-2006 (or "PUCV-2006", by its initials in Spanish) Corpus is a collection of 491 texts, totaling nearly 60 million words, which are used in the university's Industrial Chemistry, Construction Engineering,

Social Work, and Psychology undergraduate degree program curricula (Venegas, 2010). These results suggest that LCW is more contextual and interactive than the scientific articles, resembling public political speeches in this regard.

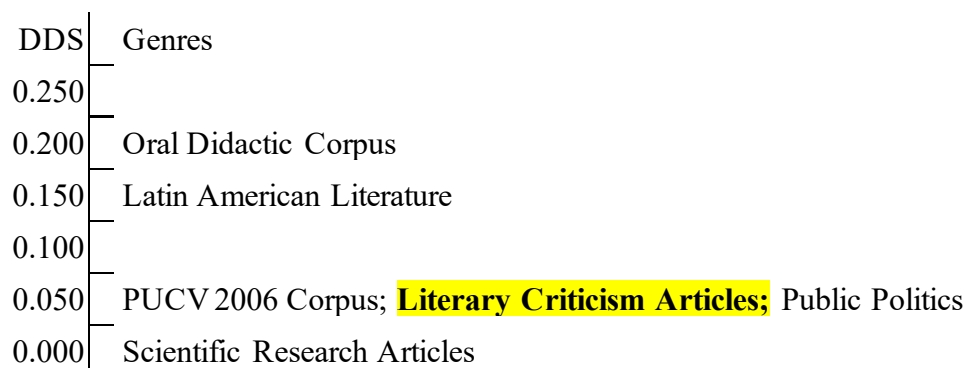


Figure 4.9. Spanish Dimension 1. Contextual and Interactive Focus

To illustrate Dimension 1 features in action, Example 26 contains a paragraph from a high-scoring Dimension 1 article.

- (26) A Pero **así** como es fácil levantar esa imagen del poeta guerrillero, también es sencillo derruirla. Hay en la propia biografía de Dalton elementos que revelan **no** tanto a un guerrillero ejemplar, absorbido por la causa, como a un militante más bien irregular, de **pronto** indisciplinado y heterodoxo. En 1967, por ejemplo, Dalton abandona el Partido Comunista salvadoreño, **debido** a "desavenencias ideológicas", y tres años más **tarde** es obligado a renunciar como miembro del Consejo Editorial de Casa de las Américas, esta vez **debido** a fricciones con Roberto Fernández Retamar y otros funcionarios de la burocracia cultural cubana. Más significativo: Dalton muere, en efecto, en la clandestinidad, mientras forma parte de la guerrilla, pero **no** a manos del ejército o de la policía sino de la guerrilla misma, cuyos dirigentes lo acusan, injusta, deleznablemente, de ser agente de la CIA. [...]. (76.txt)
- [But **just as** it is easy to raise that image of the guerrilla poet, it is also easy to destroy it. There are elements in Dalton's own biography that reveal **not** so much an exemplary guerrilla, absorbed by the cause, as a rather irregular militant, **suddenly** undisciplined and heterodox. In 1967, for example, Dalton left the Salvadoran Communist Party, **due** to "ideological disagreements", and three years **later** he was forced to resign as a member of the Casa de las Américas Editorial

Board, this time **due** to friction with Roberto Fernández Retamar and other officials of the Cuban cultural bureaucracy. More significant: Dalton dies, in effect, in hiding, while he is part of the guerilla, but **not** at the hands of the army or the police, but of the guerilla itself, whose leaders accuse him, unjustly, unjustifiably, of being an agent of the CIA. [...]

Paragraphs such as those in Example 26 appear to contain a larger concentration of cause and effect adverbial clauses features, time adverbs, and negation adverbs. These features in tandem attempt to provide signposts in the development of a line of logic in a way that involves the attention of the reader.

As a complementary illustration of Dimension 1 phenomena manifest in Spanish LCW, Example 27 contains a passage from a medium Dimension 1 scoring article with Dimension 1 features appearing in bold.

- (27) En *La nuit de l'erreur* **no** sólo se ofrece una historia legendaria, sino también la narración de los orígenes de dicha historia y, **aún** más, de los posibles desarrollos de la misma, a partir de la multiplicación de unas voces narrativas que la cuentan para que **no** muera, pues, como sentencia Dahmane, “une histoire qu'on ne raconte plus est une histoire morte” (84). Aflora **así** desde las primeras páginas un metadiscurso de la creación literaria que va cobrando peso a medida que avanza la novela, cuyo germen primero se enraiza desde el principio en la propia ciudad. Es ésta, pues, la que se identifica con la imagen de una mujer tan deseada como temida, víctima y verdugo a un tiempo; una ficción que se encarna bajo el nombre de Zina: (121).] (116.txt)

[In *La nuit de l'erreur* **not** only is a legendary story offered, but also the narration of the origins of said history and, **even** more, of the possible developments thereof, from the multiplication of some narrative voices that tell it onward so that it **doesn't** die, since, as a sentence, Dahmane <> (84). Thus, a meta-discourse on literary creation emerges from the first pages, which gains weight as the novel progresses, the root of which first takes root from the beginning in the city itself. It is this, then, that is identified with the image of a woman as desired as feared, victim and executioner at the same time; a fiction that is embodied under the name of Zina: (121).]

Example 1 contains two negation adverbs, a time adverb, and a token of *así* (“so, therefore”), indicating a mild amount of contextual involvement.

Finally, it should also be noted that there are plenty of paragraphs with very few or no Dimension 1 features, as shown in Example 28.

- (28) Es importante resaltar el carácter retrógrado que reviste la añoranza de ese signo heráldico, en un momento histórico en que **ya** se había producido la remoción de los escudos de armas en la república chilena. (3.txt)
 [It is important to highlight the retrograde nature of the longing for that heraldic sign, at a historical moment in which the removal of the coats of arms had **already** taken place in the Chilean republic.]

Despite LCW scoring higher (DDS = 0.59) than scientific articles (DDS = 0.14) on Dimension 1, these differences are to be taken in realistic terms. These values both pale in comparison to Dimension 1 scores for Latin American Literature (DDS = .163) and spoken classroom lectures (DDS = .215). LCA readers may expect an average of one more Dimension 1 feature per paragraph than in a scientific research article paragraph. However, paragraph-level frequency counts (cf. Dascalu et al., 2015) of dimension features are warranted to test this hypothesis formally. Dimension 1 features' occurring more often in LCW detail may occur, as seen in the register analysis in the previous section, as a consequence of LC articles' preference for prose over tables. Conversely, in one sample of biology articles, 80% were found to contain tables (Gray, 2015), which generally aid in involving the reader toward an understanding of the information (Wainer, 1992). The Spanish LCW corpus' favoring summaries and explanations over tables also mirrors past findings for Spanish-language humanities writing. Parodi (2010) found that, among readings used in literature doctoral programs, visual elements of any kind occurred at a rate of 0.08 per page, in contrast to biotechnology, which boasted an average of 2.00 per page.

4.3.2 Dimension 2: Narrative Focus

Dimension 2, employed by Venegas (2010) is calculated based on the Spanish imperfect aspect, first & second-person pronouns, third-person pronouns, and the periphrastic future. It has been shown to measure levels of past tense narrative in the realm to which the text refers (Parodi, 2007). The genre of text most strongly associated with this dimension in Venegas (2010) was, unsurprisingly, Latin American literature. Figure 4.10 shows how LCW scores on this dimension in relation to other genres.

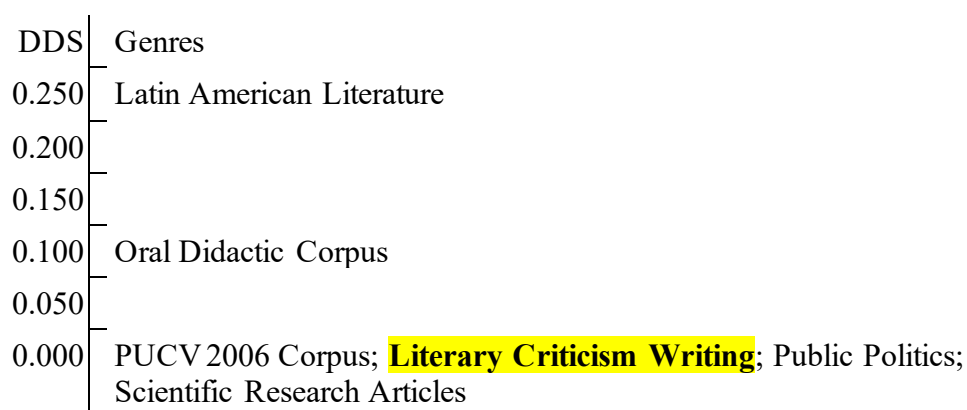


Figure 4.10. Spanish Dimension 2: Narrative Focus.

A Dimension Density Score of or close to 0.000 in Venegas' (2010) approach suggests that linguistic features of narrative writing are absent or nearly absent from a text. Despite LCW's collective fixation on novels which are often written in the past tense, Dimension 2 was the scale on which LCW scored lowest. Example 29, chosen from a medium-scoring document on this dimension, shows an attestation of this phenomenon, with just a single token of the periphrastic future while summarizing plot details. The remaining tokens of Dimension 2 features consist of the pluperfect tense.

- (29) En el co-texto del hipotexto René ha escrito a Amélie diciendo que se **va a** suicidar y Amélie acude a verlo inmediatamente. René advierte: <. La dulzura de la infancia desaparece para dar lugar al sentimiento que domina al final de la vida: el dolor. También la última remisión se vuelve más imprecisa por haber sido truncada de su contexto. René alude al hecho de que Amélie se **había** ocupado de él para evitar que se suicidara hasta que decidió tomar el velo: <. El contexto de suicidio y de religiosidad está ausente en Carpentier. Ambos protagonistas morirán por la ‘buena’ causa, o por lo menos desaparecerán en la rebeldía. Esteban no tiene fe y Sofía ya se **había** liberado del yugo de las monjas clarisas al inicio de la novela. (48.txt)
- [In the co-text of the hypotext ,René has written to Amélie, saying that he **is going to** commit suicide and Amélie goes to see him immediately. René warns: <. The sweetness of childhood disappears to give rise to the feeling that dominates at the end of life: pain. Also, the last remission becomes more imprecise for having been truncated from its context. René alludes to the fact that Amélie **had** taken care of him to prevent him from committing suicide until he decided to take the veil: <. The context of suicide and religiosity is absent in Carpentier. Both protagonists will die for the ‘good’ cause, or at least disappear in rebellion. Esteban has no faith and Sofía **had** already freed herself from the yoke of the Poor Clares nuns at the beginning of the novel.]

That the pluperfect tense co-occurs with the periphrastic future in Example 29 necessitates an analysis through the lens of speaker time (Svalberg, 1995), or the idea that a transmitter of temporal information may shift between several points in time, independently of the point in time occupied by either the transmitter, receiver, or both. Figure 4.11 contains a tense-tracking diagram (cf. Svalberg, 1995) of the shifts in verb tense from Example 29.

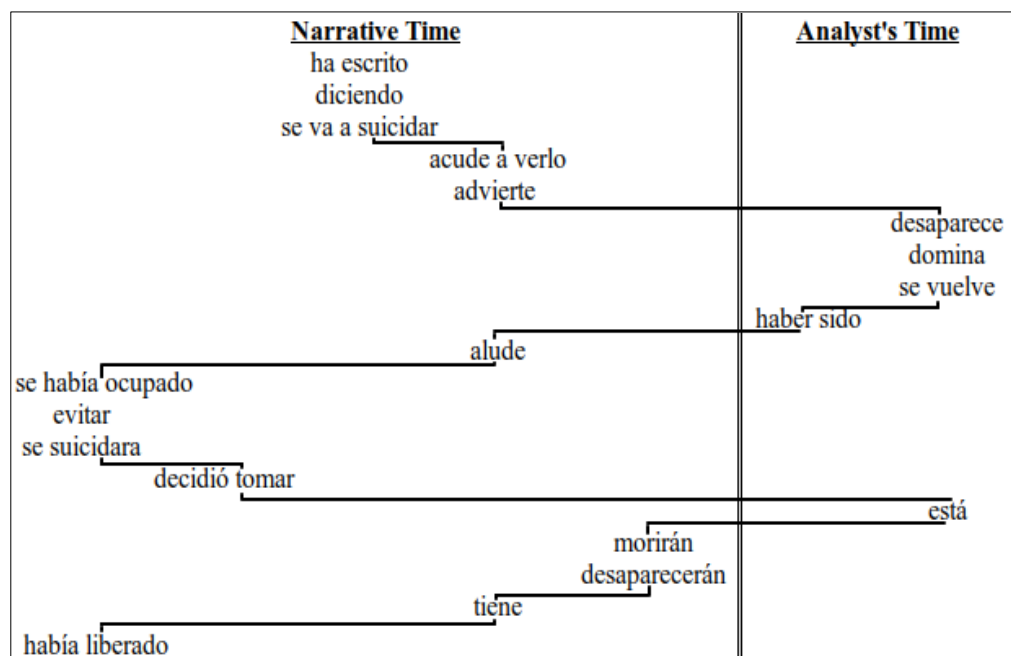


Figure 4.11. Shifts in speaker time from Example 29.

Similar to our findings for narrative linguistic features analyzed in English-language corpus in Section 4.2.3, it is proposed that the idea of four shifts in perspective articulated in Example 4 may be divided between points known as “Analyst’s Time”, or the time regarded as a baseline present from the perspective of the literary critic, and “Narrative Time”, referring to the timeline of events in the novel studied.

Example 30 contains a paragraph from a high-scoring Dimension 2 article.

- (30) *El aprendiz de brujo* conecta, en este sentido con *El profesor inútil* precisamente como la recuperación y la reorganización de los capítulos añadidos en la edición de 1934, la cual, según Ródenas y el propio Soguero **señalaban** en las suyas, no **había** acabado de cuajar 4 . Más que la reescritura de una novela, Jarnés **reescribía** su posibilidad. En *El río fiel*, uno de los relatos de *El profesor inútil*, el narrador **afirmaba** que <> (Jarnés 1999: 188)5 . Esta práctica aquí sólo tematizada es la que impulsa a Jarnés no sólo a revisar esta segunda edición o a reutilizar materiales anteriores en sus novelas del exilio sino a levantar el edificio de su obra toda. La versión de 1937 de *Locura y muerte* de Nadie **acababa** con un

"Remate y prelude" que **dejaba** abierta una novela que, en su edición de 1929, no **dejaba** espacio para ningún tipo de liberación. Por su parte, Escenas junto a la muerte se **iniciaba** con una reelaboración intensa y extensa de "Una papeleta", ahora bajo el título de "Juno", que **había** sido el último relato de la primera edición de El profesor inútil. Y, como **decíamos**, los capítulos añadidos de ésta, **formaban** el capítulo central, de título homónimo, de El aprendiz de brujo. (136.txt)

[*The sorcerer's apprentice* connects, in this sense, with *The Useless Professor* precisely as the recovery and reorganization of the chapters added in the 1934 edition, which, according to Ródenas and Soguero himself **pointed** out in theirs, **had** not quite curdled. More than the rewriting of a novel, Jarnés **rewrote** its possibility. In *The faithful river*, one of the stories of *The useless professor*, the narrator **affirmed** that \diamond (Jarnés 1999: 188) 5. This practice thematized only here is what prompts Jarnés not only to review this second edition or to reuse previous materials in his novels of exile, but to build the building from his entire work. The 1937 version of *Madness and Death of Nobody* **ended** with a "Remate and Prelude" that **left** open a novel that, in its 1929 edition, **left** no room for any kind of release. For its part, *Scenes with Death* **began** with an intense and extensive reworking of "A ballot", now under the title "Juno", which **had** been the last story in the first edition of *The Useless Professor*. And, as we **said**, the added chapters of this one, **formed** the central chapter, of the same name, of *The Sorcerer's Apprentice*.]

.]

Figure 4.12 contains a tense-tracking diagram of the verbs used in Example 30. 11 imperfect aspect constructions occur therein.

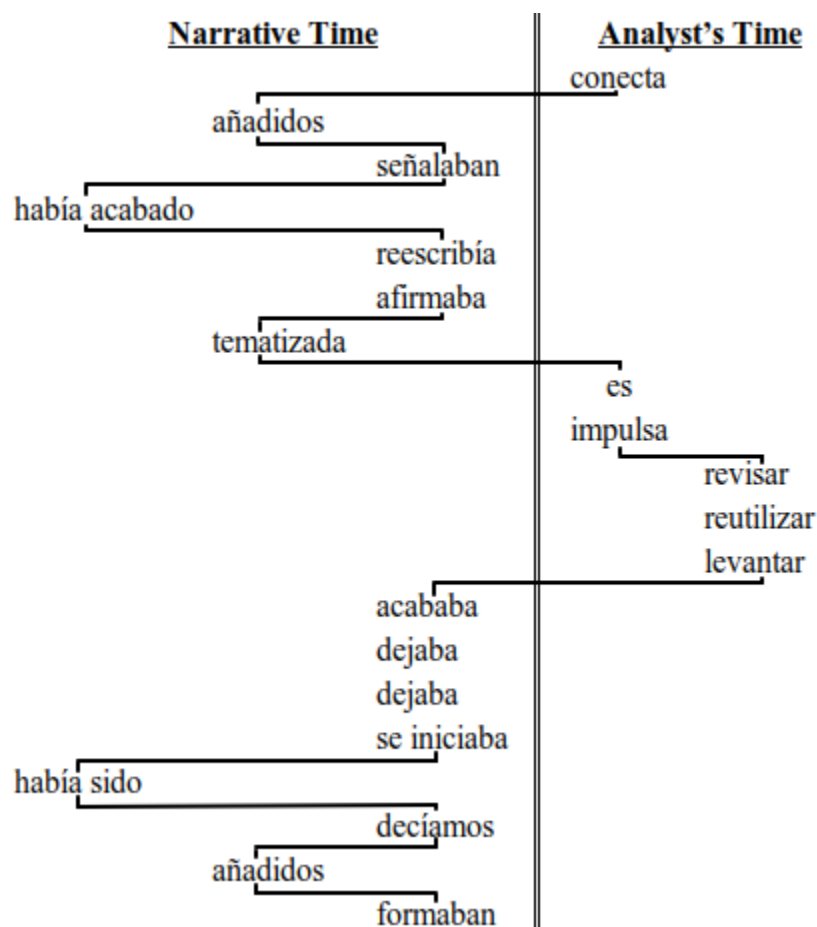


Figure 4.12. Tense tracking diagram of verb tokens in Example 30.

Parodi (2005) writes that the imperfect tense is an integral feature that serves to describe the “mundo ficticio” [fictional world] (p. 122) contained within a given text. Biber (1988) found in English that humanities prose can vary greatly in stylistic choices for the presentation of evidence. There is evidence to attest that similar levels of variation could occur in Spanish as well. The intra-corpus variation in Dimension 2 scores for Spanish ranges in Z score from -1.5 to +3.57, meaning that texts’ Dimension 2 scores varied from less than 1 standard deviation below the mean to almost 4 standard deviations above. As an example of this assertion, Example 31

shows a low-scoring Dimension 2 passage, drawing a stark contrast to the quantity of imperfect tokens in Example 30

- (31) Pues aunque allí la palabra “naturalismo” pueda aludir en efecto a una cierta forma de determinismo sociológico, o incluso “racial” (conforme a la acepción decimonónica de este término), en su comentario Payno asocia al naturalismo con una forma de representación que hace de la exhibición del horror y de lo terrible uno de **sus** mecanismos expresivos fundamentales, y que a él parece resultarle más bien ajena. De modo que en esas líneas donde explica a **sus** lectores “más timoratos” que su novela no transgredirá los límites de la decencia y de las conveniencias sociales, lo que en realidad hace este autor es deslindarse de la exhibición del horror en tanto mecanismo de representación. [...]. (92.txt)
 [So although the word “naturalism” may indeed refer there to a certain form of sociological, or even “racial” determinism (according to the nineteenth-century meaning of this term), in his commentary, Payno associates naturalism with a form of representation that makes of the exhibition of horror and that which is terrible, one of **its** fundamental expressive mechanisms, and which seems to be rather foreign to him. Thus, in those lines where he explains to **his (pl.)** “most squeamish” readers that his novel will not transgress the limits of decency and social convenience, what this author actually does is to distance himself from the exhibition of horror as a mechanism of representation.]

It appears that both high and low Dimension 2: Narrative Focus scores may hold important clues toward the communicative purpose of a given paragraph. That is, Example 31’s relative absence of Dimension 2 features appears to invest effort in making a number of inferences, rather than retelling of plot details. Example 30, however, with higher levels of Dimension 2 features, contains more context from the novel studied. Nonetheless, that such variation can occur appears to be a discipline-specific phenomenon. Biber et al. (2006) found that, generally speaking, Spanish-language academic prose utilized the imperfect tense to a lesser extent compared to other registers in a general corpus of Spanish. That some flexibility in the utilization of verb tenses is allowed, as seen in Examples 29-31, indicates potential for variation in the use of Dimension 2 features by each individual literary scholar, just as Biber (1988) found

with English texts. Additionally, Asención-Delaney & Collentine (2011), studying Spanish graduate program texts, found that the imperfect tense occurred significantly more in narrative texts than in expository texts.

4.3.3 Dimension 3: Commitment Focus

Dimension 3 measures the cumulative effects of first-person singular verb endings, private verbs, and the preterit tense. Parodi (2005) remarks that these features work together to signal that the transmitter of a discourse segment is explicit about expressing her or his attitude toward the veracity of a set of events. For example, literary works may score high on Dimension 3 due to a general preference in fiction writing for the “reliable narrator” (which, of course, has exceptions) (Booth, 2010), who is perceived as narrating and describing the events of a novel in a way perceived as consistent and accurate in the context of a fictional universe, often in the past tense. Figure 4.13 shows that LCW scored lower than scientific research articles on Dimension 3.

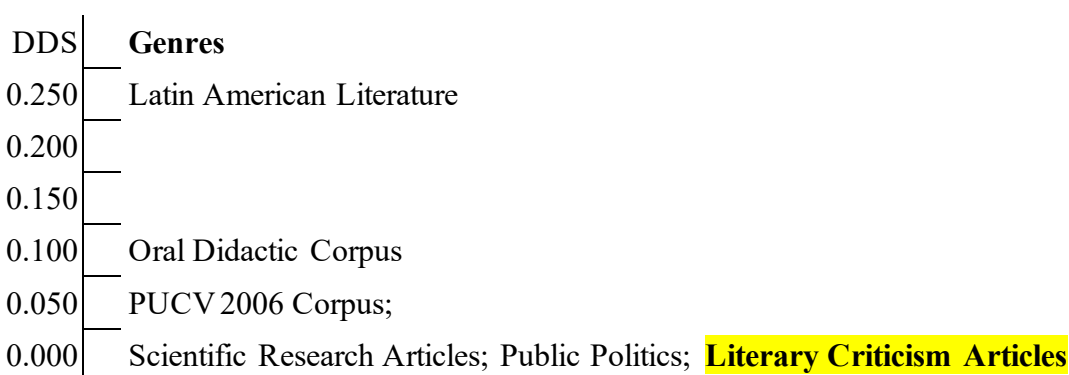


Figure 4.13. Spanish Dimension 3: Commitment focus arrangement of registers.

Example 32, a medium-scoring Dimension 3 passage, contains examples of the preterit tense and a private verb conjugated in the present.

- (32) El fragmento citado, tal como explica Magny en relación con la narrativa de Dashiell Hammet, <> (Magny, 1972: 48). En esto **consistió** precisamente la innovación narrativa que la novela negra **importó** directamente del film noir (Magny, 1972: 46-47; Simsolo, 2007: 50): voces narradoras de escenas, como la de este fragmento o la que **encontramos** al comienzo de la novela. Se trata, por tanto, de un artificio que, si bien permite la adscripción del texto de Valenzuela dentro de la tradición de la novela negra, proviene del cine. Tal sustitución de la omnisciencia por <> (Simsolo, 2007: 50), supone mucho más que la asunción de una perspectiva, también la de una consciencia narradora y fundamentalmente un tono, <> como explica Ricardo Piglia (Alí, 2007: 2). Se trata de las más sutiles cualidades de tono y atmósfera a las que se refiere Shrader. (117.txt)
 [The quoted fragment, as Magny explains in relation to Dashiell Hammet's narrative, (Magny, 1972: 48). This **was** precisely the narrative innovation that the noir fiction **imported** directly from the film noir (Magny, 1972: 46-47; Simsolo, 2007: 50): narrative voices of scenes, like the one in this fragment or the one we **find** at the beginning of the novel. It is, therefore, an artifice that, although it allows the ascription of Valenzuela's text within the tradition of the black novel, comes from the cinema. Such replacement of omniscience by "a narrator who usually takes the place of the camera" (Simsolo, 2007: 50), supposes much more than the assumption of a perspective, also that of a narrative consciousness and fundamentally a tone, "that does not it is the style, it is a relation of the one who narrates with the story; it can be a passionate relationship, it can be ironic, elegiac, distant ", as Ricardo Piglia explains (Alí, 2007: 2). These are the subtlest qualities of tone and atmosphere to which Shrader refers.]

In Example 32, the private verb *encontramos* ("we find") and its usage of the verb phrase "en esto consistió precisamente" ("in this consisted precisely"), followed by an interpretation as a completed action in history as a matter of fact "que la novela negra importó directamente del film noir" ("that the *novela negra* imported directly from film noir") seems to reflect writer's certainty expressed toward the purported historical accuracy of a claim. One also sees more variation in verb tense. As such, it was deemed expedient to summarize the variation in verb tense using another tense-tracking diagram (Svalberg, 1995), which is shown in Figure 4.14.

<u>Narrative Time</u>	<u>Analyst's Time</u>
	explica
consistió importó	
	encontramos trata permite proviene supone explica trata refiere

Figure 4.14. Tense-tracking diagram of Example 32.

Figure 4.14 suggests that some literary critics may vary in types of verb tenses used to a lesser extent than others both according to stylistic choices and communicative purpose. That is, in Example 32, tokens of the preterit contribute elements of historical context. The present tense of the verb *explicar* (“to explain”) occurs when the author cites another scholar. The present tense occurs in making an assertion about the origins of a literary technique: “Se trata, por tanto, de un artificio que, [...] proviene del cine” (“It is, therefore, an artifice that [...] comes from cinema”).

Example 33 is taken from the highest scoring Dimension 3 article. It contains a passage which shows how a literary critic came in contact with a rare manuscript.

- (33) La novela de Arkadij Maslow que aquí nos ocupa ha sido un hallazgo por varios motivos: hasta ahora era prácticamente desconocida la faceta literaria de Maslow, quien tenía cierto renombre como político de la izquierda en la República de Weimar, pero no como novelista 5 . Cuando **tuve** noticia de que este texto se hallaba inédito en una de las bibliotecas de manuscritos raros de la Universidad de Harvard, **decidí** recuperarlo sin saber de antemano cuál era su argumento ni cuáles habían sido las tesis que defendía el autor ni el enfoque histórico que en él pretendía transmitir. **Descubrí** su alcance mientras **fui** mecanografiando los 554 folios del original que, en papel cebolla, se **encuentra** en muy mal estado, lleno de rasgaduras y tachaduras, por lo que no se permite fotocopiarlo. (114.txt)

[Arkadij Maslow's novel that we are dealing with here has been a discovery for several reasons: until now, the literary facet of Maslow was practically unknown, who had a certain reputation as a left-wing politician in the Weimar Republic, but not as a novelist 5. When I **learned** that this text was unpublished in one of the rare manuscript libraries of Harvard University, I **decided** to retrieve it without knowing in advance what its argument was or what the thesis defended by the author or the historical approach in which that he intended to transmit. I **discovered** its scope while I **was** typing the 554 pages of the original, which **finds** itself in very poor condition on onion paper, full of tears and scratches, so it is not allowed to photocopy it.]

Following suit with the previous sections, the use of past tense is placed into a tense-tracking diagram (cf. Svalberg, 1995) in Figure 4.15.

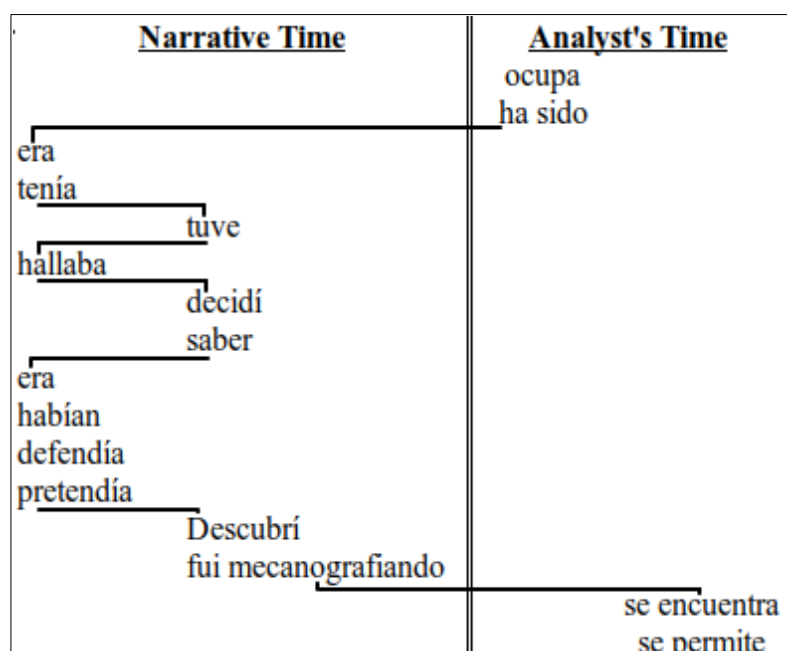


Figure 4.15. Tense-tracking diagram of Example 33.

The two private verbs (*decidir* [decide] and *descubrir* [discover]) are conjugated in the first-person preterit. As words that meet three Dimension 3 criteria, “decidí” (I decided) and “descubrí” (I discovered) added +3 each to the frequency counts of Dimension 3 features in this

particular LCA. These verbs express the certainty and decisiveness with which the researcher recounts her or his experiences with a recently-located manuscript.

In contrast, of Dimension 3 features, Example 34, taken from the lowest-scoring Dimension 3 article, contains only a single, infinitive, private verb.

- (34) La complejidad que supone la coexistencia ambivalente de pensamiento (dicotomía masculino/femenino) en la construcción de realidades debe **reconocerse** no solo en el ámbito psicológico del personaje, sino también en el lenguaje utilizado para expresar esa complejidad. (85.txt)

[The complexity of the ambivalent coexistence of thought (male / female dichotomy) in the construction of realities must be **recognized** not only in the psychological field of the character, but also in the language used to express that complexity.]

The mean Dimension 3 scores for Scientific Articles was (0.032) whereas that for LCAs was (0.024). This difference may occur partly due to the articles in Venegas' sample using higher levels of the preterit tense. Otañi (2005) has identified the preterit as a central construction used in Spanish-language methodology sections of both biomedical and linguistics research articles. One biomedical article in the GRIAL corpus (Parodi, 2005), for example, contains 73 tokens of the preterit tense, whereas the mean number of preterit tokens in the Spanish-language LCW corpus was 23. Lake & Cortes (in press) found evidence of a tendency to use formulaic phrases in the preterit in methods sections of quantitative, Spanish-language psychology articles. As was found to be the case in English, the lack of past tense narratives describing research methods may contribute to LCW earning lower Dimension 3 scores than scientific texts. Biber et al. (2006) also suggested that the use of the preterit tends to be higher in Spanish-language academic articles than in essays/newspaper columns. Thus, it is possible that on Dimension 3, LCW may alternate to some extent between norms reported for scientific writing and essay writing. Further, concerning the fact that LCAs have lower Dimension 3 scores than other

genres, Parodi (2005) writes that, relative to conversations and novels, academic texts favor information rather than the specific interlocutors or agents, opting for a more impersonal strategy focused on a dense presentation of information. It appears that the same can be argued to occur in LCW with its relative dearth of first-person declensions which appear to be the exception rather than the rule in this register. High levels of scores on this dimension do appear to contain overt commitment to the idea of veracity perceived toward an event having occurred in the past, at least in the passages chosen for illustration. However, an absence of the Dimension 3 features may not necessarily rule out the possibility of commitment to the ideas either, as we have seen in the stylistic variations in present and past tenses.

4.3.4 Dimension 4: Modalizing Focus

Dimension 4, which signals the degree to which active forms of the verb *ser* (“to be”), hedging verbs, modal verbs of possibility, predicative adjectives, and foremost, adverbs of manner express how something is said, versus something that is merely said. In Parodi’s (2005) words, “se destaca la visión particular del locutor respecto al contenido” (“the individual vision of the speaker regarding the content is highlighted”) (p. 123). Figure 4.16 shows how LCAs scored in relation to other genres.

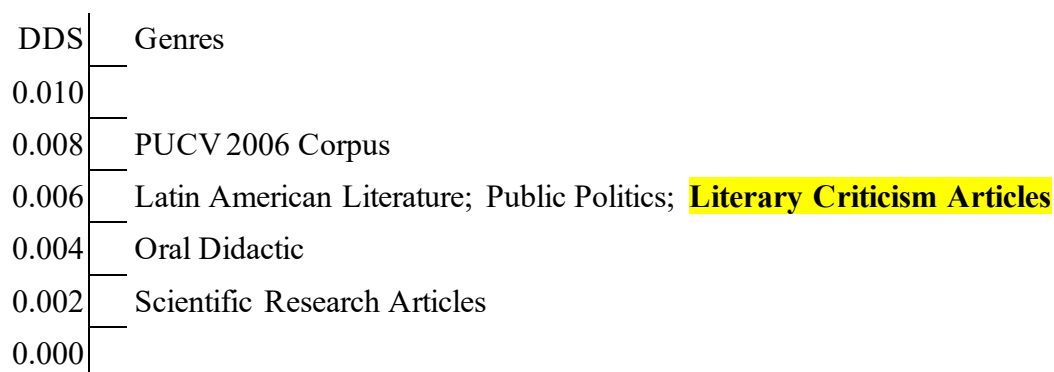


Figure 4.16. Spanish Dimension 4: Modalizing Focus.

Example 35, a medium-scoring Dimension 4 passage, appears to illustrate Parodi's (2005) assertions of clarifying the way in which the thesis is stated. The token of the conditional tense “argumento que sería la clave de lectura de mi propuesta” (“an argument that would be the key to the reading of my proposal”), harmonizes with the modalising devices “no quedan necesariamente saturados” (do not become necessarily saturated) and “en este caso me detendré únicamente” (“on this case I will center myself”). The adverb *necesariamente* [necessarily] functions as a means of qualifying the extent to which the proposition is reasonable. A subsequent nominal clause uses the conditional tense to qualify the extent of the argument.

- (35) Situando, entonces, dos alternativas más o menos polares, pero motivadas por cierta carga sublime (positiva o negativa) de la historia y la revolución, estudiaré *Trágame tierra* como novela que problematiza lo nacional y cuyos planteamientos no quedan **necesariamente** saturados por el evento revolucionario de 1979, argumento que **sería** la clave de lectura de mi propuesta. [...] En este caso me detendré **únicamente** en el asunto de su afiliación caribeña y algunas capas contradictorias de esta eventual afiliación, sobre todo porque son leídas por la novela (al menos potencialmente) desde el cuerpo dominante de lo nacional. (79.txt)

[Placing, then, two more or less polar alternatives, but motivated by a certain sublime charge (positive or negative) of history and revolution, I will study *Trágame tierra* as a novel that problematizes the national and whose approaches are not **necessarily** saturated by the revolutionary event of 1979, an argument that

would be the key to reading my proposal. [...] In this case, I will focus **solely** on the issue of their Caribbean affiliation and some contradictory layers of this eventual affiliation, especially since they are read by the novel (at least potentially) from the dominant body of the national.]

In Example 36, from a high-scoring Dimension 4 article, the two tokens of the active form of *ser* appear as part of the main clauses followed by complement clauses, the second of which comes after the modal verb of possibility *poder* (“can/may”). This sequence produces an impression of distance from the idea itself, while still asserting it.

- (36) También **es** significativo que se insista en la materialidad del libro, ya que después Carlos va en busca de los tomos, que encuentra en la habitación de Sofía. El recurso de la presencia ‘física’ de una obra **es** otro procedimiento muy carpenteriano: suele reservarlo para libros importantes. Lo aplica al Capital de Marx en La consagración de la primavera, por ejemplo. Para gran sorpresa de Carlos El Genio del Cristianismo incluye René. De ello se **puede** deducir que **es** una edición de entre 1802 y 1805, cuando este texto fundacional del romanticismo francés aún no había sido publicado por separado, modalidad en la que lo conoce Carlos de las ediciones ulteriores de La Habana.

[It **is** also significant that the materiality of the book is emphasized since afterwards, Carlos goes in search of the volumes, which he finds in Sofia’s room. A resource in the form of the ‘physical’ presence of a work **is** another very Carpenterian procedure: it is usually reserved for important books. He applies it to Marx’s *Capital* in *The Rite of Spring*, for example. To Carlos’ great surprise, *The Genius of Christianity* includes René. From this it **can** be deduced that it **is** an edition between 1802 and 1805, when this foundational text of French romanticism had not yet been published separately, a modality in which Carlos knows it from subsequent editions of Havana.]

Here, the verb *poder* [to be able] functions as a hedge for the subsequent verb *ser* [to be] to make a definitive claim about the year of publication of the novel studied.

Example 37 contains a low-scoring Dimension 4 passage.

- (37) La relación que se establece entre el acto de comer y las actividades amorosas **es** el juego entre la necesidad y el deseo entre una necesidad de procrear para sobrevivir, igual que comer. [...] **Inevitablemente** en una “fisiología del gusto” se ha de hablar del vínculo que une el lenguaje y la gastronomía: < > .
[The relationship that is established between the act of eating and amorous activities **is** the game between need and desire between a need to procreate in

order to survive, just like eating. On an anthropological and social background, the text reveals itself to us as the adventure of desire. Barthes says: <>.

Inevitably, in a “physiology of taste” one must speak of the link that unites language and gastronomy: <>.]

In discussing that LCAs (Mean = 0.056) scored higher than scientific articles (Mean = 0.032) on Dimension 4, we propose that LCW’s reliance on evidentiary presentation methods outside of tables plays a part. There was a Spearman’s Rho correlation of 0.55, $p > 0.001$ between Factor 1 and Factor 4 in the present corpus. Indeed, Parodi (2005) draws a connection between boosters and intensifiers measured on Dimension 4 and their function to influence the listener (cf. Narbona, 2000). Instead, LCW’s Dimension 4 scores are more comparable to those derived from works of literature, perhaps as a consequence of the emotive and imaginative qualities of the novels analyzed. It also bears mentioning that Beke and Bolivar (2009) found that philosophy writing was higher in modals of possibility compared to psychology and education writing. Just as the register analysis revealed several features common to English-language philosophy and LCW, there appears to be support for the idea of similarities between the same two disciplines in Spanish.

4.3.5 Dimension 5: Informational Focus

Finally, Dimension 5 is a calculation of the extent to which high levels of abstract information are condensed into as few words as possible (Parodi, 2005). Dimension 5 scores measure variation in levels of modal verbs of obligation, subjunctive mood, participles in adjectival functions, nominalizations, and prepositional phrases as noun complements. Figure 4.17 shows how LCAs scored in relation to other genres.

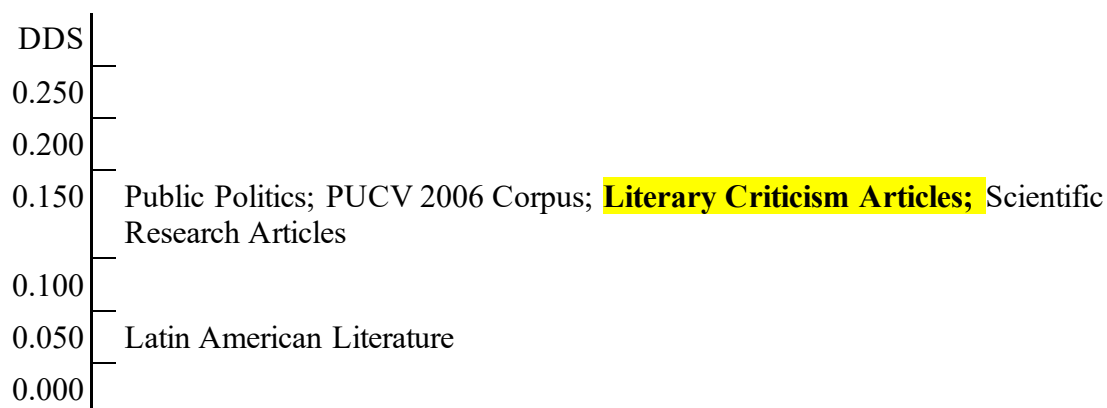


Figure 4.17. Spanish Dimension 5: Information focus

In the case of informational density, Example 38, from a medium scoring Dimension 5 article, shows vast amounts of information presented into single sentences.

- (38) Es decir, la zona gris, en confluencia a acercarse a la complejidad de los hechos **sucedidos**, puede dar luz a las memorias subterráneas y **acalladas** o que han estado encubiertas por la hegemonía de los discursos dominantes, que a veces hacen de abuso de los usos intencionales del olvido en consonancia con sus políticas ideológicas. Pero también, nos permite comprender por qué y cómo pasó lo que pasó. Me refiero a que el tejido que compone un suceso en el que participaron [millones de personas] no puede reducirse a una teoría de los dos demonios o de cuántos **sean** imaginables. Además, el **advenimiento** de gran parte de las voces de los sujetos sociales que participaron en el **encumbramiento** de los hechos, introduce en el contexto histórico actual a la **concepción** de memoria ejemplar (Jelin 2000: 58), en el que habla la comunidad en su conjunto —es decir, “nosotros” y no sólo las “voces **autorizadas**” o las víctimas de represión y abuso— para guiar a la misma en las acciones futuras, quiero decir, en lo que no **debe** volverse a hacer. En este sentido, la voz no sólo de una hija **de** exiliados sino también de una mujer, en conjunto con los diferentes actores del mundo de la novela como los [abuelos de Laura] o los vecinos —las miradas que amenazan— conforman el conjunto complejo de esas **matizaciones**. (173.txt)

[That is to say, the gray area, in confluence to approach the complexity of the events that **occurred**, can shed light on the memories that are underground and silenced or that have been covered up by the hegemony of the dominant discourses, which sometimes abuse the customs intentional oblivion in line with their ideological policies. But also, it allows us to understand why and how what happened. I mean that the fabric of an event in which [millions of people] participated cannot be reduced to a theory of the two demons or of how many are

imaginable. In addition, the advent of a large part of the voices of the social subjects who participated in the exaltation of the facts, introduces in the current historical context the conception of exemplary memory (Jelin 2000: 58), in which the community speaks in its together - that is, “us” and not just the “authorized voices” or the victims of repression and abuse - to guide it in future actions, I mean, in what should not be done again. In this sense, the voice not only of a **daughter of exiles** but also of a woman, together with the different actors in the world of the novel such as [Laura’s grandparents] or the neighbors —the threatening glances— make up the complex ensemble of those *nuances*.]

Example 39, from a high-scoring Dimension 5 article, contains high quantities of noun phrases with prepositional phrase complements, i.e., “historias reales con individuos comunes” [“true stories with common individuals”], instrumento de unificación social” (“instrument of social unification”) sentido de inclusion [“sense of inclusion”], etc.

- (39) **[Punto de encuentro] [de colectividades]**, los aparatos massmediáticos van a funcionar además como **instrumento de unificación social**, 21 cuestión que la novela reiteradamente plantea: (216–17). La progresiva **identificación** de la imagen con la idea **unido** a la teoría de que y todo ello en conexión directa con el mercado, hará que <>. 22 La **unificación** social que basada en un sentido superficial y artificial de **inclusión** opera la **televisión** en particular da lugar a una uniformidad que previene el ejercicio de libre **asociación** y **participación** social e incrementa además la brecha entre **integración** simbólica—en este caso a través de los mass media—e **integración** material—vía **redistribución** de los beneficios del **crecimiento**. 23 Ese **sentido de inclusión** operado por los aparatos massmediáticos vendrá dado desde la **televisión** a través de discursos y relatos dotados de cierta base real, discursos que . 24 Desarticulado un sentido de , 25 el tipo de historias detrás de esas imágenes remitirá ahora a microrrelatos en torno a la esfera privada y a lo individual. En su estudio , Hugo Lewin, siguiendo a Gérard Imbert, defiende que . 27 Microrrelatos que proporcionen cierta conciencia colectiva pero no en un nivel ideológico, para un sujeto que se **identifique** con **historias reales con individuos comunes**, historias . 28 En este sentido, Realidad enfatiza el hecho de que el público **tenga** un mayor interés en el plano afectivo que en el cognoscitivo, en cómo viven otros los sucesos que en los sucesos en sí. Esto se confirma al final de la historia, cuando los medios se centran en cómo vivieron los chicos las horas del secuestro, que ni siquiera tuvieron **conocimiento** de ello, en vez de en los detalles del asalto. La obra también insiste en el **grado de ficcionalización** inyectado a espectáculos con una base real, como es el caso de los reality shows, y cómo la **manipulación** que de los participantes llevan a cabo los terroristas no es más que una **continuación** del **proceso de caracterización** y, por tanto, de **ficcionalización**, propio de este **tipo**

de programas: (26). La base real del programa constituirá así el ‘**gancho**’ de **realidad** necesario para que el telespectador **acepte un pacto de credibilidad** que al mismo tiempo se **aúne** a la **ilusión** de verse representado, y por tanto incluido, a través de uno de estos ‘tipos’ humanos y que al mismo tiempo **sirva** de divertimento al ser susceptible de crear mayor **tensión** dramática. (32.txt)

[[**Meeting point**] [**of collectivities**], the mass media apparatuses will also function as an **instrument of social unification**, 21 a question that the novel repeatedly raises: (216–17). The progressive **identification** of the image with the idea **united** with the theory that and all this in direct **connection** with the market, will make <>. 22 The social **unification** that **television** operates in particular based on a superficial and artificial sense of **inclusion**, gives rise to a uniformity that prevents the exercise of free **association** and social **participation** and also increases the gap between symbolic **integration** — in this case through the mass media — and material **integration** — via **redistribution** of the benefits of **growth**. 23 That **sense of inclusion** operated by the mass media apparatuses will come from television through discourses and stories endowed with a certain real base, discourses that. 24 Disarticulated a sense of, 25 the type of stories behind those images will now refer to micro-stories around the private sphere and the individual. In his study, Hugo Lewin, following Gérard Imbert, defends that. 27 Micro stories that provide a certain collective consciousness but not on an ideological level, for a subject who **identifies** with **real stories with common individuals**, stories. 28 In this sense, Reality emphasizes the fact that the public **may have** a greater interest in the affective plane than in the cognitive one, in how other events live than in the events themselves. This is confirmed at the end of the story, when the media focuses on how the boys lived the hours of the kidnapping, who had no **knowledge** of it, instead of the details of the assault. The work also insists on the **degree of fictionalization** injected into shows with a real basis, as is the case with reality shows, and how the **manipulation** that terrorists carry out of the participants is nothing more than a **continuation** of the **process of characterization** and, therefore, **fictionalization**, typical of this **type of program:** (26). The real basis of the program will thus constitute the ‘**hook**’ of **reality** necessary for the viewer to **accept a credibility pact** that at the same time **joins** the **illusion** of being represented, and therefore included, through one of these ‘types’ human and at the same time **serve** as entertainment as it is capable of creating more dramatic **tension**.]

In contrast, Example 40 comes from a low scoring Dimension 5 article:

- (40) Aquel niño **mimado** pasó a ser de la noche a la mañana un huérfano por partida doble (primero de padre y madre y luego de los abuelos maternos). En el internado en el que le atormentaban con conceptos religiosos que le condenaban al infierno por cualquier tontería, tuvo que empezar a inventarse una historia buscando **justificación** a la partida, a su estadia en el internado y a la **falta de palabra** o presencia de quienes le querían. Despojado de guía, él mismo se convierte en su protector y objeto protegido. Empieza a creer que en el planear

una venganza y llevarla a cabo en algún momento en el futuro está la **solución** a ese **cúmulo de experiencias** sin sentido que vivió desde que fue arrancado del seno familiar. Es entonces en este punto donde cabe completar las **reflexiones de Boym** que afirma, (8). (34.txt)

[That **spoiled** child became a double orphan overnight (first of father and mother and then of maternal grandparents). In the boarding school where he was tormented with religious concepts that condemned him to hell for any nonsense, he had to start making up a story seeking **justification** for his departure, his stay at the boarding school and the **lack of word** or presence of those who loved him. Stripped of guidance, he himself becomes his protector and protected object. He begins to believe that in planning revenge and carrying it out at some point in the future is the **solution** to that **accumulation of meaningless** experiences that he lived since he was torn from the family. It is then at this point where it is possible to complete the **reflections of Boym** that he affirms, (8).]

Within the present chapter's methodological parameters, LCAs and Scientific research articles scored at comparable levels. These findings complement those of Álvarez López (2013), who discovered relatively even levels of nominalizations in Spanish-language literary criticism and economics articles. All other Dimension scores 1, 2, and 4 were significantly, negatively correlated with Dimension 5. That is, contextual, narrative, and modalizing features decreased as informational features increased in each article. These results offer further support for Parodi's (2005) remarks that Dimension 5 features favor a description of abstract detail over elaborated expansions on information using elements of interpersonal communication

4.4 Common threads in Dimension Scores in English and Spanish

Chapter 4 has discussed the results of two MDAs of literary criticism writing in English and Spanish. The Spanish Dimension 2: *Narrative Focus* is comparable to English Dimension 2 *Narrative vs. Non-Narrative Concerns*. LCW's position on both scales reveals that LCW resembles scientific writing much more than it does fiction. Similarly, nonetheless, both scales also reveal that LCW may contain some narrative features at rates slightly higher than many

scientific registers. In the same vein, given the shared concentrations of nominalizations, it is also possible to make a comparison between the English-language Dimension 3: Explicit vs. Situation-dependent reference and Dimension 5: Informational Focus in Spanish. In both languages, literary criticism and academic prose were more dense in nominalizations than were works of literature.

These findings may inform the development of reading and writing exercises for novice literary critics. Specifically, they provide examples for authentic materials that could activate novices' abilities to notice how linguistic features such as verb complement clauses, shifts in past tense, nominalizations, and qualifications of statements with adverbs are used in both languages in the presentation of evidence followed or preceded by argumentation in LCW.

4.5 Conclusion

To conclude, we will discuss the main findings of the major sections in this chapter. The register analysis in Section 4.1 revealed that literary criticism shares some characteristics of academic research articles in qualitative history and theoretical philosophy. These findings provide support for past decisions to group literary criticism, philosophy, and history writing under the umbrella of humanities writing, as did the compilers of the London-Oslo-Bergen (LOB) Corpus (Johansson et al., 1978), from which Biber (1988) later generated some dimension scores. These findings open up new possibilities in the relevance of extant linguistic research on history (Cortes, 2008; Fuentes Cortes, 2012) and philosophy (Gray, 2015) writing for improved understandings of LCW. However, that other profound differences were found between these three registers further highlights the need for a greater understanding of LCW using more detailed linguistic metrics.

Section 4.2 discussed a summary of the results of the English-language MDA. Section 4.1.1 examined the extent to which LCW resembles academic prose more than fiction writing due in part to its high quantity of nouns per sentence. Section 4.1.3 proposed that Dimension 2, in LCW, measures both the use of narrative past tense as well as the deployment of integrated citation (Hyland, 1999) strategies. Section 4.1.3 suggested that when Dimension 3 scores vary in LCW, it may occur as a result of alternating between narrating plot details and elaborating on theoretical constructs. Section 4.1.4 showed that LCW seldom expresses persuasion overtly. However, Sections 4.1.5 and 4.1.6 may, at times, play the role of (covert) expression of persuasion. Taken together, these findings suggest that English LCW resembles academic prose in its comparable levels of nominalizations, high preposition, adjective, noun, and passive construction tokens, as well as in its lower levels of past tense tokens. It regards to its utilization of complement clauses, LCW more closely resembles religious texts than academic prose. Finally, while LCW in English is less overtly persuasive than academic prose, it may possess alternative means of persuading the reader via passive voice constructions.

Section 4.3 has demonstrated the utility of Venegas' (2010) Spanish MDA approach for explaining some distinctive features of literary criticism writing in Spanish. Section 4.3 showed that LCW differs from both literary and scientific writing. Section 4.3.1 showed that, regarding context-dependency, LCW differs from scientific writing in its higher inclusion of such features as adverbial clauses, adverbs of negation, and time. Section 4.3.2 discussed how a mild degree of periphrastic future, imperfect past constructions, and personal pronouns result in LCW's more closely resembling scientific writing than literary writing. Section 4.3.3 examined the minimal levels of volitional and private verbs, first person singular inflections, and the preterit tense.

Specifically, these features contribute to implicit levels of commitment to ideas rather than overt levels thereof in a way at times consistent with other genres of Spanish-language scientific writing. Section 4.3.4 showed how LCW appears to contain fairly regular instances of hedging verbs, the verb *ser* [*be*], adverbs of manner, modal verbs of possibility, and predicative adjectives in a manner consistent with works of literature and academic philosophy articles. Section 4.3.4. showed how levels of modal verbs of obligation, subjunctive mood, prepositional phrases as noun complements, and participles in adjectival functions with nominalization to convey information at rates similarly dense like those found in other scientific research articles. In all, Spanish literary criticism writing has been shown to be neither wholly literary nor wholly scientific, but rather situated at times with norms of one end and the other. Specifically, in relation to cause & effect, negation, and adverb constructions, narrative features via the past tense, the expression of commitment through the preterit, and the qualification of statements with modalization strategies and adverbs, Spanish LCW resembles political speeches. However, in terms of informational density with terminology and suffix-derived nouns, LCW more closely resembles academic prose. , Finally, Section 4.4 considered some cross-linguistic similarities of LCW revealed through both MDAs, namely, that there is a correspondence between Dimension 2 in the Biber (1988) and Dimension 2 in the Parodi (2005) and Venegas (2010) MDA frameworks for assessing narrative features in past tense constructions. Additionally, Dimension 3 in the English framework (Biber, 1988) and Dimension 5 in the Spanish framework (Venegas, 2010) resemble each other with regards to their ability to measure dense academic prose containing suffix-derived nouns.

5 LEXICAL BUNDLES IN ENGLISH AND SPANISH LITERARY CRITICISM

WRITING

The present chapter speaks to Research Question 2, which examines the kinds of lexical bundles that occur in literary criticism, including those equivalent in structure and function in English and Spanish, as well as bundles which are language-specific. The chapter begins with a presentation of the lexical bundles in both languages chosen for the present study. It then presents the results of the functional and structural LB analyses. Afterwards, it offers comments on the equivalent and language-specific bundles organized by functional category (Cf. Biber, Conrad & Cortes, 2003; Tracy-Ventura et al., 2007) The chapter concludes with a summary of the findings.

5.1 RQ2: Lexical Bundles obtained

As discussed in Section 3.3 of Chapter 3, after the frequency and range criteria espoused in previous studies (Biber, Conrad & Cortes, 2003; Tracy-Ventura, Cortes & Biber, 2007; Cortes 2008) were specified, the resulting lists of lexical bundles were generated via AntConc. Overlapping bundles were then combined. For example, the four-word bundles *in the context of* and *the context of the* had 101 and 33 tokens respectively. However, since they had $\frac{3}{4}$ words in common, the frequency of *in the context of the* was inspected, revealing a frequency of 24 tokens. The process of their viability of existing separately is summarized in Table 5.1.

Table 5.1. Overlapping LB combination process

Variable	Step 1	Step 2
Four-word bundle A	F1	(F1 - F3)
Four-word bundle B	F2	(F2 - F3)
Potential 5-word bundle	F3	(F3)

Where F1 = the frequency of Four-Word Bundle A in a bundle pair up for analysis, F2 = the frequency of Four-Word Bundle B, F3 = the frequency of a superordinate 5-word bundle.

Table 5.2 shows the Overlapping Bundle Combination Process applied to the pairs *in the context of*, *the context of the*, and the viability of *in the context of the*.

Table 5.2. Viability test for two four-word bundles and one five-word bundle

	Step 1	Step 2	Result
<i>in the context of</i>	101	(101 - 24) =	77
<i>the context of the</i>	33	(33 - 24) =	9
<i>in the context of the</i>	24	24 =	24

Subtracting the frequency of *in the context of the* (24) from *in the context of* results in a difference of 77, indicating that *in the context of* is viable on its own. The frequency of *the context of the* (33) minus *in the context of the* (24) results in difference of 9. As the cutoff frequency for four-word bundles is 20 times per million words, *the context of the* is discarded from the final lexical bundle set, leaving only *in the context of* and *in the context of the*.

Table 5.3 contains the total unique bundle types obtained from both corpora divided by their status as four-, five-, six-, and 7+-word LBs. Table 5.3 also shows the frequency and range cutoff thresholds in line with previous analyses (Cortes, 2008; Biber et al., 1999; Cortes, 2013) for bundles of four words and longer.

Table 5.3. Percentage of bundles divided by number of words from which they are comprised

Type	EN	ES	Freq. Cutoff	Range Cutoff
4-word bundles	63.7%	61.7%	20 pmw*	5
5-word bundles	24.5%	26.7%	10 pmw	
6-word bundles	9.8%	9.6%	5 pmw	
7+ word bundles	2.0%	2.1%		
Total	102	240		

*per million words

Appendices F-H contains the resulting lexical bundles in tables divided by language and the number of words in each bundle.

The resulting list of lexical bundles was checked with previous lists of Lexical Bundles (Biber, Conrad & Cortes, 2003; Tracy-Ventura, Cortes & Biber, 2007; Cortes 2008; Pérez-Llantada, 2014; Lake & Cortes, 2020), Table 5.4 shows the number of bundles that appear to be new after checking with prior work.

Table 5.4. 4-word bundles attested in 5 studies and considered new in the present study

	English	Spanish
Attested	40%	60%
New	60%	40%
Total	63	145

The lists of potentially discipline-specific LBs are available in Appendices N & O. Examples in Spanish include *en la novela de* [in the novel of], *de la muerte de(l)*, [of/from the death of (the)], *the image of the*, and *draws attention to the*. Tentatively, it is evident that the discipline's LBs concern themselves with presenting plot events such as deaths taking places in the novels, the kinds of images involved, and interpretations thereof. We return to these functions in Section 5.3 with the functional analysis. We turn first to the structural analysis in the following section.

5.2 Structural analysis of lexical bundles

Following Cortes' (2008) approach to grouping lexical bundles, the parts of speech were identified and grouped into superordinate categories, including those of more than four words.

Table 5.5 displays the structural taxonomy divided by language and structural category.

Table 5.5. Structural distribution of LBs in English and Spanish LCW

<i>Language</i>	<i>English</i>	<i>Spanish</i>
Noun Phrases	22 (22%)	55 (23%)
NP + NP	0	3
NP + PP	20	42
NP + PP + that cl-	0	8
NP + that cl-	1	0
NP + VP	0	1
NP	1	1
Prepositional Phrases	69 (68%)	149 (62%)
PP + NP	51	130
PP + PP	12	1
PP + RB	0	1
PP + that -cl	2	11
PP + VP	4	6
That clauses	1 (1%)	7 (3%)
Verb Phrases	9 (9%)	16 (7%)
Wh-clause	1 (1%)	-
Adverbial phrase	-	13 (5%)
Total lexical bundles	102	240

Although Spanish contains more than twice the lexical bundles detected in English, the proportion of the kinds of structural classes of lexical bundles is comparable between the two languages. Similar to Cortes' (2008) findings for LBs in academic history writing in English and Spanish, the vast majority were prepositional phrase bundles, i.e., *con el objetivo de* [with the goal of], *in the face of*, and *al final de la* [at the end of]. Table 5.6 displays a list of prepositions found in the bundles of both languages in the present study.

Table 5.6. Prepositions found in LBs by language

Prepositions found in English bundles	Prepositions found in Spanish bundles
as, at, by, despite, from, in, it, of, on, over, through, to, toward(s),	a [to], con[with], de[of/from], desde[since/from], en[in], hasta[until], por[by]

Whereas 13 prepositions occurred in the English set, only 7 occurred in the Spanish set. These results support of the idea that academic writing in English has a larger quantity of prepositions, whereas Spanish may rely more on the polysemy of some prepositions in order to express similar concepts. As such, these bundles may represent a means of calling to learners' attention the differences between the two languages (Cortes, 2008).

Following in frequency were noun phrase bundles, such as *la imagen de la* [the image of the] or *the fact that the*. Verb-phrase bundles and that-clause bundles did occur (although fewer than 10% of the bundles in both languages were VP-based), aligning with past findings for history writing whose majority of lexical bundles were found to be phrasal (Cortes, 2008).

5.3 Functional analysis of lexical bundles

As discussed in Section 5.1., a substantial portion of potentially hitherto undiscovered LBs were revealed in the two corpora. In order to classify their functions, patterns observable in previously classified bundles (e.g., Biber, Conrad & Cortes, 2003; Tracy-Ventura, Cortes & Biber, 2007; Cortes, 2008) were applied to the present set. For example, taking into account the fact that abstract nouns grouped within or above prepositional phrases, such as *in the absence of* or *the nature of the* tend to be classified as “Intangible Framing Attributes”, LCW-specific bundles such as *in an interview with*, *the figure of the*, *la búsqueda de la* [the search/quest for the], and *en la descripción de* [in the description of] were classified in the same category. Similarly, noting that past functional studies tend to group bundles containing a copula (*it is important to*) and/or a mention to a noun of certainty (*the fact that the*, etc.) as stance bundles, LBs such as *despite the fact that* and *de la novela es* [of/from the novel is] were grouped into the

stance category. Explanations of the criteria for each functional category are available in Table 3.13 of the methodology chapter. Tables 5.7 and 5.8 contain example sentences in each functional category. In each example, the LB appearing therein is bolded.

Table 5.7. Examples of English LB functional categories

Category	English
Deixis	11 Toni Morrison, <i>Jazz</i> (New York: Knopf, 1992), 124. Further references are to this edition and will be cited parenthetically in the text as J. (368.txt)
Intangible Framing Attributes	In terms of his relationships with women, his aesthetic and sentimental individuality is asserted at the expense of the happiness of others. (45.txt)
Multi-Function	Destination Biafra's 'Women's war' chapter is characterised by illustrations of female solidarity and collective motherhood as Debbie and the band of Igbo women struggle to protect and feed the children in their care in the midst of the civil war. (409.txt)
Place-Institution	In the sections that follow, I touch briefly on dominant representations of black women in the United States and outline how these continue to inform social policies and ways of framing poor black women as undeserving of compassion. (471.txt)
Quantifiers	It was on this occasion that Ezeulu, the spiritual head who should have led the fight against the rest of the community, chose the path of truth and honesty on account of his devotion to the god Ulu. (279.txt)
Stance Epistemic Impersonal	Moreover, they draw attention to the fact that the liberation struggle is multidimensional. (357.txt)
Stance Obligation / Directive	It is important to mind the gap between social scripts and the actual terms of employment or immigration, especially as care should be taken not to force what Nicole Constable calls "unwarranted blurs" among these three categories of gendered migration ("Brides" 1). (349.txt)
Tangible Framing Attributes	The so-called war on terror, which has become one of the many perverse outcomes of 9/11, has not only reified geopolitical frontiers in the form of intensified border security worldwide, but has also generated new biopolitical borders in the form of a tightened governance of migrant populations and their bodies (Kuntsman and Miyake). (426.txt)
Time-Event	In the wake of the US Supreme Court's decision in <i>Plessy v. Ferguson</i> (1896) in particular, the legal status of white identity as something that held property value—a value that could be easily misrecognized or even misplaced through miscegenation—assumed a prominent place in the US cultural imaginary. (462.txt)
Topic Elaboration / Clarification	It is significant that she uses her jacket to wipe the car clean, as a symbol of the distance and sense of dignity as decorum that informed her relationship with Solomon. (561.txt)

Table 5.7 Continued

Topic Introduction / Focus	They add that Africa is one of the best sites for the study of gender identity construction, because of the particularly patriarchal nature of traditional African society and the changes it is now undergoing. (353.txt)
----------------------------	---

Table 5.8. Examples of Spanish LB functional categories

Category	Spanish	[Translation]
Intangible Framing Attributes	La idea de una nueva tierra, poblada por gentes inciviles, bárbaras e inhumanas sentará las bases del discurso colonizador de una Europa que comienza a verse a sí misma como centro del mundo, gracias a la emergencia de un nuevo espacio otro, de una nueva periferia. (73.txt)	The idea of a new land, populated by uncivilized, barbarous and inhuman people, will lay the foundations for the colonizing discourse of a Europe that begins to see itself as the center of the world, thanks to the emergence of a new space, a new periphery .
Multi-Function	Teniendo en cuenta esto, podemos afirmar que no importa tanto la referencialidad del desierto de Arizona, sino el modo en que se construye ese espacio en la escritura. (194.txt)	With this in mind, we can affirm that the referentiality of the Arizona desert does not matter so much as the way in which that space is constructed in writing.
Place-Institution	Este artículo comprende a la Ciudad de México de fines del siglo XIX como un imaginario sonoro. (43.txt)	This article understands Mexico City at the end of the 19th century as a sound imaginary.
Quantifiers	Mientras el conocimiento y la certeza, si quiera respecto a la violencia, son positivos en la mayoría de las novelas de Mendoza,14 en esta el conocimiento y la certeza dan como resultado un conocimiento negativo. (106.txt)	While knowledge and certainty, even regarding violence, are positive in most of Mendoza's novels, 14 in this knowledge and certainty result in negative knowledge
Stance Epistemic Impersonal	No cabe duda de que con esta imagen o con otras, como con la paradoja de la cinta de Moëbius, que magistralmente evoca el narrador, nos hallamos ante un nuevo tipo de laberinto, que se suma a la tipología establecida por Umberto Eco en "L'antiporfirio": (145.txt)	There is no doubt that with this image or with others, as with the paradox of the Moëbius strip, which the narrator masterfully evokes, we find ourselves before a new type of labyrinth, which adds to the typology established by Umberto Eco in "L'antiporfirio ":
Structural Only	Es evidente la posición incómoda en la que se encuentra el personaje cuando se presenta a sí mismo y las épocas de bonanza y desintegración que recuerda a lo largo del texto. (34.txt)	It is evident the uncomfortable position in which the character finds himself when he introduces himself and the times of prosperity and disintegration that he recalls throughout the text.

Table 5.8 Continued

Tangible Framing Attributes	Las canciones ayudan a Raba a expresar sus sentimientos, a aliviar su dolor, también ayudan a Molina a seducir a Valentín en El beso de la mujer araña , y ayudan a Valentín a hablar de su vida amorosa. (44.txt)	The songs help Raba express her feelings, ease her pain, also help Molina seduce Valentín in The Kiss of the Spider Woman , and help Valentine talk about his love life.
Time-Event	sólo en el momento de la muerte, cuando siente el cuchillo penetrándole en el pecho, el paquete cae al suelo, y queda liberada de su peso. (152.txt)	only at the moment of death, when she feels the knife penetrating her chest, does the package fall to the ground, and she is released from her weight.
Topic Elaboration Clarification	Como observa Infante-Voelker (2003: 52), cuando Rafael le expresa al arriero su enojo por la hipocresía que ve a su alrededor, como una justificación para abandonar el judaísmo y por lo tanto a Dios, el arriero no le da a conocer la carta de Miriam. (27.txt)	As Infante-Voelker (2003: 52) observes, when Rafael expresses to the muleteer his anger at the hypocrisy he sees around him, as a justification for abandoning Judaism and therefore God, the muleteer does not make Miriam's letter known.
Topic Introduction / Focus	En términos de una visión ginocéntrica, la lectura que ofrece Andrews se puede clarificar al considerar la marginalización de esta poeta, quien puntualiza los efectos de la guerra en torno a la mujer. (36.txt)	In terms of a gynocentric vision, Andrews' reading can be clarified by considering the marginalization of this poet, who points out the effects of the war on women.

One caveat in the above classifications is that it has been attested in past studies that some bundles' functions may change according to the context (Tracy-Ventura, Cortes & Biber, 2007). *In the midst of*, for example, occurs in the English corpus referring to the midst of segment of discourse ("in the midst of a winding complaint"), historical events ("after 9/11 and in the midst of George W. Bush's campaign for a second term"), as well as the physical-spatial sense ("in the midst of Kelly's intimidating presence"). Similarly, the Topic Introduction/Focus bundle *is one of the*, categorized as one which "signif[ies] that a new topic is being introduced" (Tracy-Ventura, Cortes & Biber, 2007, p. 226), performs not only said function but may also be argued to express stance, as in "writing that anger **is one of the** most damning traits". In cases where

potential for ambiguity became evident to the researcher, manual inspections of context sentences were carried out. Following standard LB functional classification (Biber et al., 2004), the most frequent function performed by the LB was chosen. Future studies, however, should analyze the distribution of functions performed by bundles in distinct contexts.

A complete list of LBs divided by functional category appear in Appendixes L & M. Appendix L contains the English-language set while Appendix M contains that of Spanish. In contrast to Cortes (2008), who found one bundle expressing stance in English- and three in Spanish-language history writing, literary criticism contains a larger quantity. These findings may be explained by the idea that literary criticism deals in both facts of the real world and “facts” that are accepted by readers to have occurred in works of fiction (Lake & Cortes, 2020). Bundles dealing in “facts” include the pair *el hecho de que* + determiner, *the fact that the, by the fact that*, and *to the fact that*. There are also verb-phrase bundles in this category expressing notable degrees of commitment which will be discussed more at length in Section 5.4.2.

One caveat that should be reported in the interpretation of these results is that the present study has combined overlapping bundles included lexical bundles of 5 words or more. Independently of any disciplinary customs of literary criticism, the combination of overlapping bundles may be one reason for which numbers in some functional categories may register as higher than those in categories found in Cortes (2008), who only studied four-word bundles.

Table 5.9 contains the percentage division of bundles by functional category and language.

Table 5.9. Distribution of LB functional category by language

	English	Spanish
Deixis	1.0%	0.0%
Intangible Framing Attributes	34.3%	42.5%
Multi-function	2.9%	8.8%
Obligation / Directive	2.0%	0.0%
Place-Institution	1.0%	0.8%
Quantity Specification	1.0%	0.4%
Referential Tangible	1.0%	0.8%
Stance Epistemic Impersonal	14.7%	5.8%
Structural Only	0.0%	4.6%
Tangible Framing Attributes	2.9%	9.6%
Time-Event	18.6%	13.3%
Topic Elaboration / Clarification	16.7%	6.7%
Topic Introduction / Focus	3.9%	6.7%
Total	102	240

Appendices L & M contain each bundle identified in both languages divided by functional category.

Just as Cortes (2008) found for academic history writing, Spanish-language articles in LCW predominate in Multi-function reference, Topic introduction, Structural, and Intangible and Tangible Framing Attributes bundles. English, however, has more Obligation/directive, stance, topic elaboration, and Time-Event bundles. That structural categories occurred only in Spanish appears to be a consequence entirely attributable to lexicogrammatical conventions. For example, Cortes (2008) observes that *en el que se*, consisting of four words, is often translated as the relativizer “which”, consisting of only a single word. Similarly, other functional categories of lexical bundles may contain echoes of this tendency. For example, the Spanish dataset has seven multifunction PP bundles containing the trigram *a partir de* [from], meaning that direct equivalents in English of some bundles containing *a partir de* [from] might only consist of two

words in English, for example, *a partir de la* vs. *from/through the*. Similarly, although the LB *por medio de un* could be literally translated as *by means of a/the*, another option used in some English-language academic prose is *via*. Similarly, *a través de* can sometimes have equivalencies in *across* and *through*. However, that the sequence is expressed in three words contributes to the occurrence of the Multi-function bundle types *a través de la/las/los*, *a través de su(s)*, and *a través de un/a*. Concerning the topic introduction bundles in Spanish, four of them contain conjugations of *tratarse de*, which is often used functionally as an equivalent of *to deal with* [a topic]. Finally, the Topic Elaboration/Clarification category in English has four bundles containing the trigram *as well as*, which can generally be achieved in Spanish with the bigram *así como*.

5.4 Equivalent analysis of lexical bundles

Laying the groundwork for the present section, with structural and functional equivalents identified, the conditions were favorable for the identification of equivalent lexical bundles in both languages. Table 5.10 lists bundles deemed equivalent, organized by functional category.

Table 5.10. Equivalent LBs by functional category

Category	Spanish	English
Referential Intangible Framing Attributes	desde el punto de vista de	from the perspective of a
	desde el punto de vista de la / las / los	
	desde la perspectiva de la(s) / los	
	desde la perspectiva de(l)	
	en el caso de	in the case of
	en el contexto de	in the context of
	en el contexto de la(s) / los	in the context of the
	en el plano de la	on the level of the
en el proceso de	in the process of	

Table 5.10 Continued

	en la medida en que	
	en la medida en que el/la	to the extent that
	en la medida en que se	
	en relación con el/la	in relation to the
	la forma en que	the way in which
		the ways in which
	la historia de la	the history of the
	la imagen de la	the image of the
	de la novela es	of the novel is
Stance Epistemic Impersonal	el hecho de que la/el/las/los	the fact that the
	no es más que	can be read as a can be seen as a
	en el hecho de que	to the fact that
	a finales del siglo xix y	
	del siglo xix y principios del	at the turn of the century
	las primeras décadas del siglo xx	
	a la vez que	at the same time as at the same time that
Table 5.10 (Continued)		
	a lo largo de la historia	over the course of the
Referential Time-Event	al final de la	by the end of the at the end of the
		toward(s) the end of the
	al final de la novela	by the end of the novel at the end of the novel
	del siglo xix y principios del	at the turn of the century
	el final de la novela	the end of the novel
	en la primera parte de	at the beginning of the in the first half of the
	en la segunda parte	in the second half of the
	la segunda mitad del siglo	the second half of the
Discourse Organizers Topic Introduction / Focus	como una de las / como uno de los	as one of the
	es una de las / es uno de los	is one of the

5.4.1 *Time-Event*

The analysis revealed a substantial amount of equivalent referential bundles referring to times and events (cf. Cortes, 2008). Examples 1-4 illustrate this tendency.

- (1) **Al final de la novela**, el protagonista se ve como el perro fiel de su pueblo abandonado, un perro que no ha podido, o sabido, portarse de otra manera. (17.txt)
[**At the end of the novel**, the protagonist is seen as the faithful dog of his abandoned town, a dog that has not been able, or known, to behave otherwise.]
- (2) **By the end of the novel**, Oskar discovers that the key was not meant for him but was mistakenly left inside the vase, which his father bought at an estate sale. (476.txt)
- (3) **En la primera parte de** Tierra adentro, Rafael muestra poca comprensión consciente de su viaje, como si estuviera fuera de sí mismo, narrándose a sí mismo como un observador. (27.txt)
[**In the first part of** Inland, Rafael shows little conscious understanding of his journey, as if he were outside himself, narrating himself as an observer.]
- (4) **In the first half of the novel**, Mann's depiction of the duchy hovers between reactionary conservatism and radical critique: [...] (Dedner). (320.txt)

Viera Echevarría (2014) proposes that a presentation of evidence comes before a logical conclusion in literary criticism conference presentations in Spanish. The high quantity of lexical bundles in both languages in the Time-Event category used to report the moment at which details in the novel occur appear to offer support for this assertion in written articles.

5.4.2 *Stance Epistemic Impersonal*

Stance epistemic impersonal bundles occurring in LCW, share syntactic and topical similarities when describing, comparing, and evaluating fictional and ideological entities. A notable point of cross-linguistic variation attested in prior research (Perales Escudero & Swales, 2011) was found in the ways in which scholars in English and Spanish differ when laying out series of claims about meanings and social critiques hidden in works of literature. It appears that

there is some support for Perales Escudero's & Swales' (2011) claims with the bundle pair *can be read as a* and *no es más que*. Examples 5 and 6 are sentences containing the former:

- (5) Similarly, Sapphire's previous claim that women must push back against violence **can be read as a** call for women to rally against gender-based violence. (471.txt)
- (6) Welcome to Hard Times **can be read as a** Western which portrays the suffering of a pioneering community due to a merciless environment and the destruction brought about by the external "Bad Man from Bodie". (281.txt)

Both examples place forth interpretations of works that make larger inferences toward social issues. Examples 7 and 8 contain example sentences of the latter.

- (7) La azarosa biografía de Zina **no es más que** la fábula segregada por una ciudad vieja de experiencias y prolífica en fantasías y supersticiones, cuya población masculina teme, sin dejar de sucumbir a ello. (116.txt)
[Zina's haphazard biography **is nothing more than** the fable segregated by an old city of experiences and prolific in fantasies and superstitions, whose male population fears, while still succumbing to it.]
- (8) La interrupción en esta primera parte, **no es más que** la experiencia transcrita de García Madero, ese aparente acento en el detalle como la única manera posible de contar la historia del real visceralismo. (13.txt)
[The interruption in this first part **is nothing more than** the transcribed experience of García Madero, that apparent emphasis on detail as the only possible way to tell the story of real visceralism.]

In the examples above, then, the bundles appear to operate in a capacity similar to that expressed in the sentences involving *can be read as a*, although in the case of Examples 7 and 8 with more commitment to the idea, thus appearing to offer support that Perales Escudero's and Swales' (2011) assertions also hold relevance for intercultural variation in Spanish- and English-literary criticism articles. However, it should still be noted that *no es más que* has only 30 tokens, while *can be read as a* and *can be seen as* combined have 66 tokens. In corpora of over 150 articles in both languages, numerous articles will contain neither of these LBs. However, that the tendency is observable in some articles in both languages warrants attention by future scholars.

5.4.3 *Discourse Organisers: Topic Elaboration/Clarification*

Tracy-Ventura, Cortes & Biber (2007) write of Topic Elaboration bundles that they “are used to add additional information about the speaker’s message” (p. 226). Examples 9 and 10 merit attention as a means of topic elaboration/clarification in their way of expanding upon assertions on the nature of irony in the case of Example 9 and on the themes of the work studied in Novel 10.

- (9) Edgar Johnson, en *A Treasury of Satire*, destaca la función de la ironía **como una de las** estratagemas más poderosas de la sátira indirecta, y la labor del ironista como un fingidor consumado en el arte de decir lo contrario de lo que dice. (3.txt)
[Edgar Johnson, in *A Treasury of Satire*, highlights the role of irony **as one of the** most powerful stratagems of indirect satire, and the ironist’s work as a consummate pretender in the art of saying the opposite of what he says.]
- (10) El título juega con el lector y juega también con su propio significado, porque, a lo largo de la obra, cuando la reflexión sobre la muerte propia empieza a perfilarse **como uno de los** temas centrales, se revela otro sentido de “años de indulgencia”, que, a esta nueva luz, vendrían a ser los años en que todavía la muerte lo perdona a uno. (157.txt)
[The title plays with the reader and also plays with its own meaning, because, throughout the work, when the reflection on own death begins to emerge **as one of the** central themes, another sense of “years of indulgence” is revealed, which, in this new light, would become the years when death still forgives you.]

In Examples 11 and 12, the functions repeat. In the case of the former, additional details about a notable aspect of a novelist are authored, In the case of the latter, the importance of a theoretical construct is specified.

- (11) And yet, although any number of authors from colonized lands in Africa and Asia have been invoked in this fashion, Rizal merits attention **as one of the** first truly globalized authors of the modern era. (322.txt)
- (12) Masculinity **as one of the** major aspects in a gender-theoretical perspective forms the theoretical framework. (353.txt)

5.4.4 *Topic Introduction / Focus*

Tracy-Ventura et al., (2007) describe Topic Introduction / Focus bundles as being utilized “to either signify that a new topic is being introduced or to direct the listeners’ or readers’ attention to a specific part of the discourse” (p. 226).

In Example 13, the bundle “es uno de los” serves to introduce a topic, while in Example 14, its role as focusing on an already-established topic appears.

- (13) Hans Reiter, un escritor admirado por un grupo de profesores de literatura, **es uno de los** protagonistas de 2666 y firma bajo el seudónimo de Benno von Archimboldi. O, las divagaciones sobre el mal y los crímenes en el testimonio del detective Abel Romero (396-97), están tomadas de las mismas ideas que este personaje desarrolla en la novela Estrella distante (1996). (13.txt)
[Hans Reiter, a writer admired by a group of literature teachers, **is one of the** protagonists of 2666 and signs under the pseudonym Benno von Archimboldi. Or, the ramblings about evil and crimes in the testimony of detective Abel Romero (396-97), are taken from the same ideas that this character develops in the novel Estrella distante (1996).]
- (14) Quizás el minotauro **es uno de los** muchos empleados como funcionarios en quien la memoria porosa y neurótica del protagonista cree escuchar uno de los autores del secuestro de su madre. (99.txt)
[Perhaps the minotaur **is one of the** many employed as civil servants in whom the protagonist’s porous and neurotic memory thinks he hears one of the perpetrators of his mother’s kidnapping.]

Indeed, in these English-language examples, we also see evidence for the sense of the topic focus usage of these bundles, Example 15 focuses on the idea of major themes from a region. Example 16 focuses on circumstances used as justification for an explanation of phenomena in the novel of study,

- (15) In his important thematic analysis of the literature of coastal BC from 1984, Pritchard argues that <> (38) but that “nature as victim of man,” **is one of the** major themes of the literature of the region (38). (533.txt)
- (16) Stephen feels a melancholy at his heart, suffering from a nostalgia which **is one of the** factors that leads him at the end to recant, to go to the US Embassy in Paris, and name names, and thus put an end to his trajectory and its logic. (63.txt)

5.4.5 Referential: Intangible Framing Attributes

Given the high quantity of referential bundles in both languages, it stands to reason, probabilistically, that some of them would be in common. Many of them also function in identical fashions, as we see in examples 17-20.

- (17) Para Freud, la idea de lo siniestro ha de estudiarse y entenderse **desde el punto de vista de** la estética filosófica, ateniéndose a las nociones de percepción o sensación, ya que estas se dedican al estudio de la percepción de realidades, teniendo en cuenta la sensación que despierta en el ser humano. (39.txt)
[For Freud, the idea of the sinister has to be studied and understood **from the point of view of** philosophical aesthetics, adhering to the notions of perception or sensation, since these are dedicated to the study of the perception of realities, taking into account the sensation that awakens in the human being.]
- (18) Yuniór reconstruye el relato a partir de lo que le han contado, **desde el punto de vista de** su padre. (160.txt)
[Yuniór reconstructs the story from what they have told him, **from the point of view of** his father.]
- (19) Manfred Jurgensen's little-known novel *The American Brother* (2007) avoids the tendentiousness that inheres in more conventional novels written **from the perspective of a** political liberalism that blindly condemns a perceived shift toward absolutism within the post-9/11 nation-state. (45.txt)
- (20) In a few lines, *Highway* clearly illustrates how English is understood (or not understood) **from the perspective of a** Cree child, and Gabriel's lack of understanding emphasizes the linguistic and cultural boundary between the boy and the Roman Catholic brother. (427.txt)

In the above cases, attributes of either theoretical or fictional entities are used to frame further remarks.

This tendency is notable also in the following bundles dealing with context which occur in Examples 21 and 22.

- (21) También cabe mencionar los referentes extraliterarios que se inscriben en la novela: la figura idealizada de la familia nuclear, la microempresa propagada durante el primer sexenio de los presidentes panistas y la fragmentación de los cárteles de la droga **en el contexto de la** guerra contra el narcotráfico, declarada al inicio del segundo sexenio: <> (Sada 2012: 86). (129.txt)

[It is also worth mentioning the extraliterary references that are inscribed in the novel: the idealized figure of the nuclear family, the micro-enterprise propagated during the first six-year term of the PAN presidents and the fragmentation of the drug cartels in **the context of the** war against drug trafficking, declared at the beginning of the second six-year term: <> (Sada 2012: 86).]

- (22) Whilst this note may on the face of it reflect Sophie's respect for Steven to the extent that she feels that she cannot go without informing him, when read **in the context of the** patriarchal culture's gendered role and familial power configurations, the note can be viewed as covertly implying a disruption of Steven's patriarchal power and familial influence. (273.txt)

Examples 23 and 24 also qualify how statements might be viewed.

- (23) **En el plano de la** práctica discursiva, la novela de Carlos Fuentes se forma y desforma en la "frontera", espacio para la enunciación, el cual condiciona la producción del discurso social en una sociedad fronteriza calcinada por las transgresiones culturales, la discriminación, las injusticias sociales, la corrupción, la lucha por la supervivencia y las diferencias de género. (112.txt)
[**At the level of** discursive practice, Carlos Fuentes' novel is formed and deformed at the "border", a space for enunciation, which conditions the production of social discourse in a border society burned by cultural transgressions, discrimination, social injustices, corruption, the fight for survival and gender differences.]
- (24) The partition of character-space among various main characters in the family reinforces **on the level of the** character-system the transgenerational vision undergirding Oscar's claim. (304.txt)

Examples 25 and 26 deal with processes and how they related to the fictional and real worlds.

- (25) A partir de los conceptos desarrollados por Malcom Parkes (1976) entra con fuerza esta idea de pensar que **en el proceso de** traducción, donde claramente incide el factor lingüístico de pasaje de una realidad lingüística a otra, también implica la traducción de un sistema editorial a otro. (182.txt)
[Based on the concepts developed by Malcom Parkes (1976), this idea of thinking enters strongly **into the process of** translation, where the linguistic factor of passage from one linguistic reality to another clearly affects, also involves the translation of an editorial system into other.]
- (26) This emphasis on affective interaction imbricates the majoritarian as well as the minoritarian subject **in the process of** racialization. (308.txt)

Also of great disciplinary specificity is the bundle pair in which images as well as archetypes can be discussed in both languages, as seen in examples 27-30.

- (27) [...] **la imagen de la** perversa mujer insaciable es una creación masculina que justifica el control, la opresión y la violencia. (163.txt)
[[...] **the image of the** perverse insatiable woman is a masculine creation that justifies control, oppression and violence.]
- (28) **La imagen de la** jaula se refiere a la modernidad: es la jaula que encierra, que permite ver lo que está adentro, pero que no permite volar. (46.txt)
[**The image of the** cage refers to modernity: it is the cage that encloses, that allows us to see what is inside, but does not allow us to fly.]
- (29) It is worthwhile, then, to focus on **the image of the** hated oppressor which, for the protagonist-narrator of the novel, is Russia (excluding the period of 1918–39) and, subsequently, the Soviet Union, both represented here as a wild and barbarian formation. (547.txt)
- (30) So all seems lost, collapsed into **the image of the** dislocated, homeless Emily. (63.txt)

5.4.6 *Summary of equivalent bundle analysis*

In sum, it appears that there is a fair attestation of disciplinary conventions in both languages to fulfill similar kinds of functions. Time event, Topic-elaboration, Topic introduction, and Intangible framing attributes function largely similarly in both languages. Students learning to produce LCW in English or Spanish as a second language will do well to mind the prepositions in some of these bundles, as in the case of the pair *en el plano de la* and *on the level of the*, or *desde el punto de vista de* and *from the point of view of*. These examples are chosen since “en” can be translated as “at”, “in”, and “on” while “desde” can be translated as both “from” and “since”.

However, in line with Cortes’ (2008) call for more parallel comparable corpus LB studies to understand “similarities and differences in patterns of language use across different academic discourse communities.” (p. 56), intercultural academic discourse community variation was found in stance bundles *can be read as a*, *can be seen as*, and *no es más que* (“is nothing but”). Nevertheless, the extent to which these differences present any obstacle specifically to academics

socialized in English-speaking LCW contexts aiming to submit to Spanish-language literary criticism journals or vice versa is not currently known and should be explored by future studies.

5.5 Language specific bundles

In the present section, we discuss language-specific lexical bundles as organized by functional category. We comment on potential, linguaculturally specific conventions that may contribute to some bundles being language-specific. To begin, Table 5.11 shows the distribution of language-specific bundles divided by functional category. Lists of language-specific bundles are found in Appendices N & O.

Table 5.11. Language specific bundles by functional class

	English	%	Spanish	%
Intangible Framing Attributes	16	16%	47	20%
Multi-function	2	2%	20	8%
Obligation	2	2%	0	-
Place Reference	1	1%	1	0.4%
Stance Epistemic Impersonal	5	5%	0	-
Structural Only	0	-	11	5%
Tangible Framing Attributes	2	2%	21	9%
Time-Event	4	4%	15	6%
Topic Elaboration / Clarification	14	14%	2	1%
Topic Introduction / Focus	0	-	11	5%
Language-specific Total	46	45%	127	53%

The table shows that, with the exception of stance, Topic Elaboration, Obligation, and Place-reference LBs, higher percentages of language-specific bundles occurred in Spanish.

Morphosyntactic antecedents for the higher quantity of framing attributes bundles in Spanish are discussed in Sections 5.5.1 and 5.5.6.

As for the differences in stance bundle types, there are several factors at play. At the surface level, *by the fact that* has a comparable counterpart in Spanish *por el hecho de que*. However, during the combination of overlapping bundles, it was found that *por el hecho de* when followed by an infinitive verb assumes the properties of a Topic Introduction/ Focus bundle, e.g., “[l]os familiares son acusados de drogadictos, alcohólicos y proxenetas, solo **por el hecho de** ser pobres,” (163.txt) [The family members are accused of being drug addicts, alcoholics, and pimps solely for their being poor]. *Por el hecho de* occurred 20 times, whereas *por el hecho de que* occurred only 10. As such, *por el hecho de* was prioritized and categorized as a Topic Introduction / Focus bundle.

The remaining differences in both Stance and Topic Elaboration bundles appear partly attributable to English’s preference for longer constructions compared to equivalent forms in Spanish. Concerning Topic Elaboration / Clarification bundles, consider the English-language bundle *as a symbol of*. It contains four words. Its equivalent in Spanish, however, is *como símbolo de* which, despite occurring 20 times, contains only 3 words. Consequently, it was not observable using the cutoff frequency, number of words, and range criteria for 4-word LBs.

In the same way, the stance bundle *it is worth noting that* has one possible equivalent construction in the Spanish corpus: the verb *valer* (to be worth) + infinitive and its more emphatic construction *vale la pena* (lit. is worth the pain/effort) + infinitive, which initiate verb complement clauses 7 times in the corpus, a number comparable to the 10 occurrences of *it is worth noting that* in the English corpus. The equivalent construction *se destaca* (is/are

emphasized), occurs 20 times. Further, the construction *es importante* + infinitive occurs 51 times, both of which come out higher than the 27 tokens of *it is important to* and the 10 of *it is important to note that*. These Spanish equivalencies perform many of the same duties in fewer words than do their English-language counterparts, obviating the Spanish constructions' appearance in lexical bundle lists based on criteria for 3-, 4-, and 5-word bundles. As such, these cases bring to light occasions in which Spanish LCW, like other Spanish genres perhaps, prefers constructions with more brevity than one finds their English-language counterparts.

5.5.1 *Intangible Framing Attributes*

Intangible framing attributes bundles prevail in both languages. Tracy-Ventura et al. (2007) describe them as working to outline quantities and attributes of nouns. However, there are more Spanish-specific bundles in this category than in English.

- (32) En *El Señor Presidente*, y en general **en la narrativa del** dictador, la escritura y el archivo están contenidos en el cuerpo del déspota. (99.txt)
[In *El Señor Presidente*, and in general in the **dictator's narrative**, the writing and the archive are contained in the body of the despot.]
- (33) En este artículo, se analiza **la representación de la** migración en la novela *Crónica de músicos y diablos* de Gregorio Martínez: (97.txt)
[In this article, **the representation of** migration in Gregorio Martínez's *Chronicle of Musicians and Devils* is analyzed:]
- (34) **The figure of the** well-assimilated and successful Vietnamese refugee not only allows for the revisionist casting of Americas role in Southeast Asia as defender and savior, but it is also appropriated as justification for present and future US military interventions overseas. (508.txt)
- (35) Mays is not suggesting that we return to live within the dangers of hidden passageways and alleys; rather, he bemoans the fact that **in the name of** commerce, the urban experience is being turned into "unending day" (151). (530.txt)

One factor at play seems to be that the usage of the Saxon genitive in academic writing is less frequent than in other registers (Grafmiller, 2014). It has apparently decreased, along with

periphrastic of genitives in the past century in scholarly writing in English in favor of premodifying nouns (Biber et al., 2016). Thus it appears that compound nouns may also play a part in the inventory of possible strategies that result in a variety of options for bundles that would mirror those in Spanish. Nonetheless, the frequency of periphrastic genitive constructions is still noticeable, as it was previously found in English (Biber, Conrad & Cortes, 2003) more than 60% of academic prose bundle types contained a noun phrase + post modifier composite. As such, it makes sense that they would prevail in the language-specific subset as well.

5.5.2 *Place Reference*

There are comparable levels of place reference also. There are comparable levels place reference, *de la Ciudad de México* [from Mexico City] and *in the united states*.

- (36) Un acaudalado criollo **de la ciudad de México** decide partir en compañía de su criado Francisquillo hacia el Viejo Continente, punto geográfico que considera como su auténtico hogar. (88.txt)
[A wealthy Creole **from Mexico City** decides to leave in company of his servant Francisquillo to the Old Continent, a geographical point that he considers his true home.]
- (37) Assimilation **in the United States**, they argue, depends on the attainment of “dominant norms and ideals,” including “whiteness, heterosexuality, middle-class family values” (“Dialogue” 344) often foreclosed for racial minorities. (468.txt)

Mexico is the most populous Latin American Spanish-speaking Country. Despite efforts to represent a variety of Anglosphere countries in the English-language corpus, literature and the affairs of the U.S. remain recurring topics in journals around the world. However, that so few place reference bundles passed the lexical bundle threshold speaks to the value of including journal articles from a wide variety of countries. One detriment of this approach, however, is that the possibility of measuring regional variation, a concern articulated in past academic writing

studies (Biber et al., 1999, Cortes, 2008). is forfeited due to writing by authors socialized in any one country not surpassing 1,000,000 words, a threshold that has recently been reaffirmed as a necessary quantity of information from which to extract LBs (Bestgen, 2019).

5.5.3 *Time-Event*

An English-specific phenomenon is found in the Time-Event category of bundles denoting that something has occurred after something deemed negative by the writer, *the face of the, in the aftermath of the, in the wake of the*, etc., which appears to be a language-specific phenomenon at least within the parameters of examining the problem from the phraseological window of 4+-word lexical bundles.

Moreover, the LBs *in the aftermath of* and *in the wake of* contain poetic, infrequent noun-phrases to describe that something terrible has transpired, with potentially more dread to follow shortly after, as in Examples 38-39.

- (38) Chesnutt's education, as Brodhead himself points out, was sponsored by the Peabody Fund, an early Northern philanthropy created by banker George Peabody **in the aftermath of** the Civil War. (374.txt)
- (39) Although Ruth and Oliver suspect that the lunchbox drifted to the island **in the wake of** the 2011 tsunami, Ruth never finds out for sure. (533.txt)

Future research should investigate the semantic prosody of these bundles (cf. Cortes & Hardy, 2013).

Specific time-event bundles to Spanish deal with linguistically specific ways of syntactically framing time periods in the cases of *de la década de(l)* [from the decade of (the)], *de la década de los* [from the decade of the (pl.)], *de la década de los cincuenta* [from the 50s], *en la década de los* [in the decade of the (pl.)] which are followed by a tens place number to

delineate the decade in which an event occurred. In contrast, in English the tens place number is pluralized in fewer words (i.e., “the 30s”) to denote the same concept.

Here also is where more ways of saying when something occurs as synonyms for when: *a la hora de* [at the time of] and *en el momento de* [at the moment of]. Finally, there are some established ways of saying *a lo largo de* [through] followed by *relato* [story], *siglo* [century], *texto* [text], *novela* [novel], *toda la novela* [the whole novel], and some rarer seven-word bundles *en la primera parte de la novela* [in the first part of the novel]. Bundles of this type may stem from the fact that syntactic genitive case constructions are not features of modern Spanish, instead achieved via prepositional circumlocution after the genitive case disappeared during its evolution from Latin (Lloyd, 1987; Pharies, 2015).

5.5.4 *Obligation-Directive*

As for the English divergences, the obligation directive majority appears to occur as a consequence of English not being a pro-drop language, meaning it requires a pronoun with verbs. As such, most impersonal directives are required to contain the pronoun "it" as in *it is important to note that*, in addition to its requirement of the preposition *to* when presenting a verb in the infinitive form.

Concerning obligation-directive bundles, there are just two in English: *it is important to* and *it is important to note that*. These are instances identified as the anticipatory it (Quirk et al. (1985). As it is used in lexical bundles, Biber et al. (2003) note that it includes stance. Subsequent studies note that important can be swapped for other adjectives e.g., interesting, in order to express the meaning of the researcher (Jalali, 2017). Biber Conrad & Cortes (2003) add

that “attitudinal bundles express speaker attitudes towards the actions or events described in the following proposition: desire, ability, obligation, or intention” (p. 81-82). These findings are also reflective of the stance indication capabilities previously highlighted as characteristic of verb complement and adjective complement clauses in academic writing (Biber, 1988). We see examples of the importance conveyed from the author to the reader using these bundles in Examples 40-43.

- (40) As such, **it is important to** debunk the myth of Canadian superiority with respect to racism in order to better consider Roth’s novel in the context within which it is written. (501.txt)
- (41) Hence **it is important to** understand that this replayed traumatic separation is also, crucially for Morrison, the primal scene of slavery: the separation of mother and child. (309.txt)
- (42) **It is important to note that** Peter Bacho's *Cebu* has been read as both an authentic representation and a social satire. (313.txt)
- (43) **It is important to note that**, in 1941, hoarding gold coins was not simply idiosyncratic or even irrational behaviour. (462.txt)

Future studies should attempt to identify equivalent rhetorical strategies in Spanish.

5.5.5 *Structural only*

In Spanish, this category of bundles consists largely of Prepositional phrases + + that-complementizers in all cases whose equivalents, as noted by Cortes (2008), are usually “which” or “in which” in English. Manifestations of this phenomenon occur in Examples 44-46.

- (44) La novela continúa insistiendo en restarle valor al cuerpo del indígena y siguen dos escenas **en las que las** mujeres indígenas son tratadas como mercancía propiedad del patrón y comparadas directamente con las vacas. (8.txt)
[The novel continues to insist on devaluing the body of the indigenous person and two scenes follow **in which the** indigenous women are treated as merchandise owned by the patron and compared directly to cows.]
- (45) Una vez que el joven observa cómo se consumen sus padres en la pira, él mismo decide tomarse la justicia por su mano urdiendo un plan **en el que se** vengaría del

atroz final de sus padres. (39.txt)

[Once the young man observes how his parents are consumed in the pyre, he himself decides to take justice into his hand, devising a plan **in which he** would take revenge on the atrocious end of his parents.]

- (46) Por ello, en sus reflexiones se mezclan y confunden las ideas propias con las ajenas, mejor o peor asimiladas y siempre tamizadas por la fecunda imaginación infantil, como sucede en el capítulo “La alegría de Dios”, **en el que el** protagonista se enfrenta a la tentación de la carne, el remordimiento y la amenaza del castigo: <>. (127.txt)

[For this reason, in his reflections his own ideas are mixed and confused with those of others, better or worse assimilated and always sifted by the fertile children’s imagination, as happens in the chapter “The joy of God”, **in which the** protagonist faces the temptation of the flesh, remorse and the threat of punishment: <>.]

5.5.6 Tangible Framing Attributes

With regards to tangible framing attributes, a high volume of bundle types in this category specific to Spanish merited a further subdivision to gain a more comprehensive view of their functions. In Spanish, we have, broadly, three groups of Tangible Framing Attributes Bundles. These are listed in Table 5.12. Examples showing characteristics of each them follow.

Table 5.12. Subgrouping of Spanish-specific tangible framing attributes bundles

Group	Bundles
1: Intra-novel focus	en la novela de, en la novela el/las/los/la, en la novela se, en el texto de, en la obra de [in the novel of, in the novel the, in the novel + deagentivization marker, in the text of, in the work(s) of]
2: Properties focus	de la novela como, de la novela de, de la novela en, de la novela histórica, de la novela que, de la novela se, de la novela y, de las novelas de(l), de la novela de la [of the novel as, of the novel of, of the novel in, of the historical novel, of the novel that, of the novel + deagentivization marker, of the novel and, of the novels of the, of the novel of the]
3: Fictional entity focus	de los personajes de, de los personajes que, de los personajes de la novela, de la casa de(l), a sí mismo(s) como [of the characters of, of the characters that, of the characters of the novel, of the house of the, to themselves as]

Group 1, Intra-novel focus, taking on the properties of framing bundles which focus in on the textual circumstances which aid in making analyses of their textually adjacent assertions. In Example 47, the Group 1 bundle, *en la obra de* [in the work(s) of] for instance, notes the qualification that it is in specifically one novelist's works that the human body contains a number of possible meanings.

- (47) El cuerpo **en la obra de** Yasunari Kawabata es signo abierto a múltiples significados. (54.txt)
[The body in Yasunari Kawabata's work is a sign open to multiple meanings.]

One possible interpretation is that the assertion following the bundle may not hold validity in other contexts.

The pattern repeats with a qualification of comparison between Dostoyevsky and Bakhtin in example 48.

- (48) Tanto **en la obra de** Dostoiévski como en la tradición con la cual la conecta Bajtin —los géneros “serio -cómicos” de la Antigüedad— se hace patente toda la cadena de renovaciones genéricas que implica la construcción de un yo auténtico. (157.txt)
[Both **in the work[s]** of Dostoevsky and in the tradition with which Bakhtin connects it - the “serious-comic” genres of Antiquity - the whole chain of generic renovations that the construction of an authentic self implies becomes evident.]

This tendency is also evident in 49 and 50.

- (49) Es lo que ocurre **con la novela de** Élmer Mendoza *Cóbraselo caro* (2005), la cual tiene su origen explícito en la obra de Juan Rulfo Pedro Páramo (1955). (106.txt)
[This is what happens **with the novel by** Élmer Mendoza *Cóbraselo caro* (2005), which has its explicit origin in the work of Juan Rulfo Pedro Páramo (1955).]
- (50) Como veremos, **en la novela se** revelan dos movimientos paralelos y en principio contradictorios: (8.txt)
[As we will see, **in the novel** two parallel and in principle contradictory movements **are** revealed:]

Group 2, properties focus, in contrast to discussing the intra-novel focus, which is usually initiated with the preposition “en” or “in”, is initiated with the preposition “de”, and as a framing

bundle, speaks to properties perceived as belonging to or characteristic of the novels studied.

Examples 51-52, for instance, present underlying assertions about novels.

- (51) Esta lógica va en contra de la idea **de la novela como** un mapa dado que presupone un cierto desarrollo. (133.txt)
[This logic goes against the idea **of the novel as** a map since it presupposes a certain development.]
- (52) Este objetivo **de la novela histórica** empieza por su propio estatuto genérico, el cual encierra una paradoja, ya que el término “novela histórica” encierra una contradicción de conceptos. (120.txt)
[This objective **of the historical novel** begins with its own generic statute, which contains a paradox, since the term “historical novel” contains a contradiction of concepts.]

Finally, Group 3, Fictional entity focus, focuses on properties of fictional entities, whether

objects or entities which form part of the fictional universe, as seen in Examples 53-56.

- (53) Esta intransigencia fue el evento disparador que llevó al joven de doce años **de la casa de** sus padres a un mundo para el que no estaba preparado. (27.txt)
[This intransigence was the trigger event that took the twelve-year-old boy **from the house of** his parents to a world he was not prepared for.]
- (54) La historia **de La casa de las Bellas Durmientes** es sencilla. (54.txt)
[The story **of The House of the Sleeping Beauties** is simple.]
- (55) El narrador se percibe **a sí mismo como** un personaje de una trama caricaturesca, de corte expresionista, que evoca la presencia de Valle-Inclán. (145.txt)
[The narrator perceives **himself as** a character in a cartoonish plot, of an expressionist style, that evokes the presence of Valle-Inclán.]
- (56) Pequod narra la relación de tres generaciones de una misma familia, a través **de los personajes del** abuelo, Manuel; el padre, Ahab y el hijo, narrador. (186.txt)
[Pequod narrates the relationship of three generations of the same family, **through the characters of the** grandfather, Manuel; the father, Ahab and the son, narrator.]

In Category 3, we see more evidence for the possibility that periphrastic genitive constructions in Spanish boost the probability of formulaic sequences within the language

crossing LB thresholds at rates greater than one finds in English, a language which may also opt for Saxon genitive constructions.

The only tangible framing reference bundles that are specific to English is: *in the form of* and *in the form of a*, illustrated in Example 57.

- (57) Herms's playful assemblage, in fact, combines both designs, since his random accumulation of salvaged books **in the form of** Libra's scales, which double as the outstretched arms of the libra/rian to whom he pays homage in his portrait, is central, but the dangling cup, to the viewer's left, serves as a reminder of the individuation of objects as well as of the warmth of Herms's own transformative spiritual vision. (472.txt)

Broad in meaning, it may at times perform some of the functions performed by Groups 1-3 in the Spanish dataset.

5.5.7 *Topic Elaboration / Clarification*

Within this category one finds those LBs previously attested in past studies, i.e., *on the one hand*, which "reflect relationships between prior and coming discourse" (Biber, Conrad & Cortes, 2003, p. 79). Examples 58-61, containing newly-attested bundles, show the connection points between prior and subsequent discourse segments.

- (58) The kitchen, with its stereotypical mark **as a symbol of** women's domesticity and inferiority, emerges as a symbolic space that, in a patriarchal sense, marginalises all who are associated with it. (273.txt)
- (59) Qangule (1974:1) interprets the river **as a symbol of** purification, redemption, spiritual rebirth, hope, new life or regeneration. (282.txt)
- (60) Gee will then need **to come to terms with the** fact that trauma, pain, and suffering are not only feelings that reside inside him, but also forms of affect that circulate and shape the spaces he occupies. (426.txt)
- (61) For Janie, the porch's public and private dimensions allow her **to come to terms with the** interior and exterior dimensions of her psyche. (316.txt)

A highly frequent Topic Elaboration bundle specific to Spanish is *y por lo tanto*:

- (62) Cualquier muestra de sentimentalidad, de emotividad se iguala a debilidad **y por lo tanto** a feminidad que, en el caso de los varones, deriva en homosexualidad, conceptualizada como degradante e indeseable. (187.txt)
 [Any sign of sentimentality, of emotionality is equated with weakness **and therefore** femininity that, in the case of men, leads to homosexuality, conceptualized as degrading and undesirable.]

Por lo que se refiere a can also be enlisted for similar duties of linking a concept with a further explanation elaboration.

- (63) **Por lo que se refiere a** la modalidad narrativa, el “relato de acontecimientos”, al igual típico de la novela romántica, hace que éstos se presenten a los lectores como sucesos distantes y difíciles de leer por su cantidad y densidad. (125.txt)
 [As for the narrative modality, the “account of events”, as is typical of the romantic novel, makes these present themselves to the readers as distant and difficult events to read due to their quantity and density.]

In Example 63, the bundle is working as topic elaboration, but may be used in other contexts as well across the corpus.

Another bundle in this category is *en la novela que nos ocupa* [“in the novel at hand”]. It appears to function as a means of elaborating on a phenomenon worthy of study described within a given novel, as seen in Example 64.

- (64) Uno de los grandes aciertos, entre otros, de Ángel de Campo en sus crónicas y, por supuesto, **en la novela que nos ocupa** es que no solo enumera, sino que adjetiva con metáforas el sonido. (43.txt)
 [One of the great successes, among others, of Ángel de Campo in his chronicles and, of course, **in the novel at hand** is that he not only enumerates, but uses sounds adjectivally with metaphors.]

5.5.8 *Topic Identification / Focus*

There is a notable attestation of Spanish-specific topic identification/focus bundles. *De la novela que nos ocupa*, for instance, appears to draw focus to the novels analyzed in Examples 65-66:

- (65) Así ocurre con la historia de Zina, personaje nuclear **de la novela que nos ocupa**, eje de una leyenda que su propia voz fomenta, tal y como reza el *excipit* de un prólogo que nos sitúa sin ambages en el terreno de la ficción. (116.txt)
 [“This is the case with the story of Zina, the central character **of the novel in question**, the axis of a legend that her own voice encourages, as the *excipit* of a prologue says, which places us unambiguously in the field of fiction”]
- (66) Pese a que no es mi intención la de hacer un cotejo exhaustivo entre una versión —la de la memorias— la otra —la **de la novela que nos ocupa**— de los hechos, porque se me antoja absurdo y porque el mayor o menor grado de semejanza y fidelidad entre las memorias y la novela no afecta en nada a la postura aquí defendida, sí que quiero citar algunos ejemplos concretos de esta coincidencia textual entre El árbol de la ciencia y las memorias de Pío Baroja para certificar la existencia dentro del conjunto de la obra del escritor de ese «espacio autobiográfico» al que ya he aludido. (155.txt)
 [“Although it is not my intention to make an exhaustive comparison between one version - that of the memoirs - the other - that **of the novel in question** - of the facts, because it seems absurd to me and because the greater or lesser degree of similarity and fidelity between the memoirs and the novel does not affect in any way the position defended here, I do want to cite some concrete examples of this textual coincidence between The Tree of Science and the memoirs of Pío Baroja to certify the existence within the set of the work of the writer of that “autobiographical space” to which I have already alluded.”]

Others in this category tend to be used for identifying the topic first, rather than elaborating on it, at least in the present set of texts as is visible in Examples 67-69.

- (67) Como se mencionó, David es un muchacho ingenuo, dueño **de una serie de** cualidades físicas especiales que su vida campesina ha acrecentado. (187.txt)
 [As mentioned, David is a naive boy, owner **of a number of** special physical qualities that his peasant life has enhanced.]
- (68) Igualmente, **de la misma manera que** en la búsqueda de la estética o de lo bello, la noción de lo ominoso interpela la capacidad sensorial, y provoca, en este caso, una sensación de angustia o de terror. (39.txt)
 [Likewise, **in the same way as** in the search for aesthetics or the beautiful, the notion of the ominous questions the sensory capacity, and in this case causes a sensation of anguish or terror.]
- (69) Entre las misivas que recibió el autor gracias a la publicación en Madrid de La luciérnaga, hubo una de F. Orozco Muñoz, testimonio **de uno de los pocos** críticos que valoraron el retrato de la oralidad del pueblo, una de las grandes virtudes de la obra. (151.txt)

[Among the letters that the author received thanks to the publication in Madrid of *La luciérnaga*, there was one by F. Orozco Muñoz, testimony **of one of the** few critics who valued the portrait of the orality of the people, one of the great virtues of the work]

5.5.9 *Stance*

Finally, Biber, Conrad & Cortes (2003) describe stance bundles as offering “a frame that is used for the interpretation of the following proposition. [...] epistemic stance bundles comment on the status of the information in the following proposition: certain, uncertain, or probably, possible.” (p. 81-82) English has more language-specific stance bundles than did Spanish. We see a strong attestation of the following stance epistemic impersonal LB working in the capacity of attitudinal remarks toward a phenomenon in examples 70-72.

- (70) This analysis of the role of women **in the novel is** by necessity sketchy; in particular, I am unable here to expand either on the binary between Ada and her husband’s Irish lover, Eva Tanguay, or on Ada’s travesty role as substitute for George right after his stroke, which points to another level of (gendered) performativity. (557.txt)
- (71) One part of the cultural problem encoded **in the novel is** that the struggle between Anglotexans and Mexicotexans has become less obvious, because U. S. nationalism has successfully domesticated the border, foreclosing criticism of the imperialist violence that made Mexicotexans U. S. subjects. (365.txt)
- (72) **It is worth noting that** there is a history of female participation in African liberation struggles – female combatants in Umkhonto we sizwe, the armed wing of the African National Congress (ANC), being a case in point. (409.txt)

Through the lens of, shown in examples 73-75, although not specifically stance in and of itself, is classified as such due to the inferential circumstances in which it predominantly appears:

- (73) The material deficiencies and human intimacies that poverty causes are seen by Mata **through the lens of** didactic western eyes. (543.txt)
- (74) The early-twentieth-century diversification of black manhood in the United States can be usefully viewed **through the lens of** the New Negro concept. (314.txt)

- (75) The objective world is always interpreted **through the lens of** Romantic aesthetics, through a tradition of poetic images and tropes that frame the outside world. (45.txt)

In these three examples, for example, the the bundle *through the lens of* is preceded by an agentless passive construction, which, as discussed in Section 4.2.6, appears to be used to make more inferential interpretations.

A dominant pattern of the sole Spanish-specific stance bundle is that it is sometimes preceded by verbs of cognition and emotion, followed by claims taken for granted as fact either by the writer or the writer studied by the writer. Like the secondary function of private verbs revealed in LCW in Section 4.1.2 of Chapter 4, the LB *en el hecho de que* may serve as a device for initiating integrated citations (Hyland, 1999). Examples 76- 78 illustrate this possibility.

- (76) La torsión de Prado a los registros históricos se aprecia **en el hecho de que** efectivamente cuando Dutrou Bornier, el comerciante de la Polinesia, ocupa la isla, contrae matrimonio con Koreto, a quien nombra reina, como un modo de proclamar a Rapa Nui protectorado francés. (132.txt)
[Prado's twist to historical records is seen **in the fact that** when Dutrou Bornier, the Polynesian merchant, occupies the island, he marries Koreto, whom he appoints queen, as a way of proclaiming Rapa Nui a French protectorate. 16.]
- (77) Es necesario insistir **en el hecho de que** el sustrato referencial de esta novela ha sido previamente ficcionalizado por el Estado mexicano en aras de desvirtuarlo o francamente desaparecerlo. (187.txt)
[It is necessary to insist **on the fact that** the referential substrate of this novel has been previously fictionalized by the Mexican State in order to distort or frankly disappear it.]
- (78) El mismo proceso de disolución del cuerpo deseante se observa **en el hecho de que** la autora convierte a su protagonista en piedra, cancelando precisamente la posibilidad de que, por encima del dolor y el fracaso, la memoria registre y retenga el cuerpo de la sensualidad y el amor-tabú.(28.txt)
[The same process of dissolution of the desiring body is observed **in the fact that** the author turns her protagonist into stone, precisely canceling the possibility that, over pain and failure, memory registers and retains the body of sensuality and love-taboo.]

5.6 Summary

The present chapter has shown how lexical bundles exhibit features of academic writing shared across languages in literary criticism as well as certain phenomena which are language specific. It first discussed the kinds of potentially discipline-specific bundles revealed in the LB analyses. Structural tendencies for bundles in both languages were compared, demonstrating that distributions thereof largely resemble each other in both languages. The functional distribution of the LBs indicated that one finds a moderate degree of comparability between languages.

Although Spanish predominated with Framing bundles and topic introduction and structural bundles, English had more obligation, stance, time-event, and topic-elaboration bundles. An analysis of the equivalent and quasi-equivalent bundles evinced a comparable degree of cross-linguistic similarities in framing, situating events in time and place, as well as topic introduction. Also of interest is that, in bundles expressing stance, the notion of greater epistemic commitment to ideas among some Spanish-speaking scholars (Perales Escudero & Swales, 2011) was attested in this corpus as well. Section 5.4 and its subsections focused on language-specific bundles, noting that English had more Stance bundles specific to the language, whereas Spanish had more framing and topic introduction ones. An additional item of discussion was that the English-specific time event bundles utilized poetic language and tended to occur in negative contexts. It is worth bearing in mind that, as Section 3.6.1 of Chapter 3 illustrated, four-word LBs constituted an average of 3.6% (SD = 1.2%) of the words per Spanish article and about 1.4% of the words (SD = 0.6%) in an English LCA. Thus, despite contributing to a minuscule amount of the prose in an LCA, the present chapter has shown that some LBs, particularly those in the Time-Event and Stance categories, appear to perform important, communicative purposes. The

next chapter will explore or present an exploratory way of linking these bundles to the communicative purposes of writers in literary criticism.

6 COMMUNICATIVE PURPOSES AND LEXICAL BUNDLES

The present chapter responds to RQ3 and RQ4: the range of communicative purposes that can be identified in a body of literary criticism articles and how they relate to lexical bundles.

The chapter begins with a summary of the communicative purposes hierarchy presented in Section 3.4 of Chapter 3. Some LB examples that may correspond to each communicative purpose are offered in reference to Research Question 3, the range of communicative purposes that can be identified. Section 6.2 explores Research Question 4 by reporting the communicative purposes linked to specific LBs. Section 6.3 presents an analysis of equivalent bundles and their communicative purposes. In Sections 6.4 and 6.5, participants' judgments are analyzed in terms of agreement and LB Saliency. Section 6.6 offers some tentative explanations for the results. Lastly, 6.7 summarizes the chapter's findings and presents directions for future research.

6.1 RQ3: Communicative Purposes Hierarchy

As mentioned in Chapter 3, in an earlier pilot study (Lake & Cortes, 2019), sentences in 8 research articles were categorized according to a taxonomy designed using Stage I of Mayring's (2014) Qualitative Content Analysis (QCA). In QCA, analysts work independently to reduce a string of text to its essence via inductive categorization techniques. Table 6.1 contains a summary of the hierarchy

Table 6.1 *Communicative purposes hierarchy*

General	Specific	Definition
Reporting	Defining	A definition of theoretical constructs or other key terms.
	Describing	A description of a person or a state of affairs on a real or imagined plane.
	Retelling	A restating of information, whether from the author's biographical details, research by other scholars, or the novel itself.
Inferring	Explaining	A presentation of information with some degree of inference not immediately obvious to the reader.
	Evaluating	Evaluations which may come across as more subjective than experts would normally allow in other disciplines.
	Comparing	Can be used to establish an interplay of similarities between: characters, elements in the plot, other scholars, or other works of fiction and the theoretical framework

Table 6.2 contains LBs, communicative purposes, and example sentences which were categorized according to my own preliminary analyses.

Table 6.2. *Lexical bundles used in sentences of each communicative purpose*

Comm. Purpose	Bundle	Sentence
Defining	on the one hand	As I mentioned in the introduction, on the one hand immunity in its modern political meaning refers to the exemption of the individual from the 'expropriating features of <i>communitas</i> ' (Esposito, <i>Bíos</i> 50, emphasis in original) – the isolation of the private sphere in order to protect and foster life. (45.txt)
	A partir de sus	La crítica ha denominado a este proceso intertextualidad, si nos atenemos al término que más recorrido ha tenido y que difundió Julia Kristeva (1978) a partir de sus estudios sobre Bajtín. (156.txt) [Criticism has called this process of intertextuality, if we stick to the term that has had the most travel and that Julia Kristeva (1978) spread from her studies on Bakhtin.]
Describing	in the United States	The novel, as yet unpublished, tracks the lives of three Palestinian women; the narrator of "The Way Back," a Palestinian woman now living in the United States , narrates a journey to upstate New York: "We passed a giant sign that said CIGARETTES in ten- foot red letters [...]". (375.txt)

Table 6.2 (Continued)

	de la casa de	La enfermedad de éste es la de mi hermano Darío. Pío describe el huerto de Burjasot y algo de la casa de la calle de Samaniego, de Valencia (Baroja, 1998: 198). (155.txt) [His illness is that of my brother Darío. Pío describes the orchard of Burjasot and something of the house on Calle de Samaniego, in Valencia (Baroja, 1998: 198).]
Retelling	in the midst of	Set in Fort de-France, Martinique, the novel recounts an investigation into the mysterious death of a Creole storyteller named Solibo Magnificent who chokes and dies in the midst of his discourse on the Savannah (a local park). (318.txt)
	de la historia que	no ha podido consolar al Ángel de la historia que lloraba por un mundo roto por su eje y les augura algo mejor que lo que él ha visto: «Tampoco habéis visto lo enrojecida de sangre que está esa esfera en algunas partes. Ni Dios quiera que la veáis nunca, y no podáis ver partido el eje del mundo, ni al Ángel desconsolado» (2013: 11). (156.txt) [He has not been able to console the Angel of history who was crying for a world broken by its axis and he predicts something better than what he has seen: «You have not seen how reddened that sphere is in some parts either. God forbid that you ever see it, and you cannot see the axis of the world broken, nor the disconsolate Angel »(2013: 11).]
Explaining	that there is a	For example, Sapphire dedicates her text “To Children Everywhere,” not only relying on the trope of the American child as the site of hope, but also implying that there is a sort of universal suffering (or potential for that suffering) among children. (471.txt)
	desde la perspectiva del	La creación de dos narradores observadores permite a Baroja guardar una distancia respecto a lo narrado y dar la versión de los acontecimientos desde la perspectiva del extraño — del extranjero, del que viene de fuera — y, como tal, tiene libertad para mezclarse y tratar de comprender las dos facciones. (156.txt) [The creation of two observer narrators allows Baroja to keep a distance from what is narrated and to give the version of events from the perspective of the stranger - the stranger, the one who comes from outside - and, as such, he is free to mix and try to understand the two factions.]
Evaluating	the fact that the	The fact that the victim’s throat is cut, or “snicked,” by the word, from the inside, is at once comic and sad. (318.txt)

Table 6.2 (Continued)

de la novela es	<p>Si entendemos al personaje autobiográfico en el sentido más ortodoxo, como un trasunto de la persona real del autor que revive las situaciones anteriormente vividas por este y adopta en la novela sus actitudes y opiniones, creo sin duda alguna que el único personaje autobiográfico de la novela es el protagonista. (155.txt)</p> <p>[If we understand the autobiographical character in the most orthodox sense, as a transcript of the author's real person who relives the situations previously lived by him and adopts his attitudes and opinions in the novel, I believe without a doubt that the only autobiographical character in the novel is the main character.]</p>
Comparing on the level of the	<p>[T]he trope of memory is linked to Euclid on the level of the narrative structure as well: "Euclid, lost to the rest of the world, was Matussem's private land, like the country his parents tried to leave as they made lives in Jordan, as they let go of their children's memories and let them grow up as Jordanians [...]". (375.txt)</p>
en el caso de	<p>La diferencia entre ambos radica en que, mientras el primero se inspira en la filosofía de Schopenhauer absorbida por Baroja, el segundo es un personaje de raíz puramente nietzscheana; de hecho, en el caso de Iturrioz, Sobejano retrocede tres años en la producción barojiana y sostiene que ya en <i>La dama errante</i> (1908) —la novela en la que Iturrioz aparece por primera vez como un personaje secundario— nos encontramos ante un personaje autobiográfico: una contrafigura del autor que entronca con el pensamiento de Nietzsche y que ejerce de «portavoz de Baroja» (Sobejano, 2004: 368-369). (155.txt)</p> <p>[The difference between the two lies in the fact that, while the former is inspired by Schopenhauer's philosophy absorbed by Baroja, the latter is a character with a purely Nietzschean roots; in fact, in the case of Iturrioz, Sobejano goes back three years in the Barojian production and maintains that already in <i>La dama errante</i> (1908) - the novel in which Iturrioz appears for the first time as a secondary character - we find ourselves before an autobiographical character : a counter-figure of the author that connects with Nietzsche's thinking and who acts as "Baroja's spokesman" (Sobejano, 2004: 368-369).]</p>

6.2 RQ4: Communicative purposes linked to lexical bundles

To examine the range of communicative purposes represented in the above data, we explore, in this section, bundles that corresponded to each communicative purpose in the hierarchy.

Defining

Example 1 contains a bundle which, while not apparently exhibiting any features specific to Defining seen in the analytical criteria (in Appendices A and B), serves the purpose all the same.

- (1) Nussbaum explains that compassion is also grounded in a sense of human flourishing **as well as an** understanding of the resources that must be present for a person to succeed. (471.txt)

The focal point for the rating in this category was the interplay between the author Nussbaum and the possibility of a specific operationalization of the word compassion, as seemingly nonspecialist words often are in LCW (Biber & Gray, 2016)

Example 2 contains a similar passage in Spanish.

- (2) La crítica ha denominado a este proceso intertextualidad, si nos atenemos al término que más recorrido ha tenido y que difundió Julia Kristeva (1978) **a partir de sus** estudios sobre Bajtín. (156.txt)
[Critics have termed this process intertextuality, if we accept this ubiquitous term that Julia Kristeva (1978) spread **from her** studies on Bakhtin.]

In Example 2, it is much clearer that the Bakhtinian construct of intertextuality is in the process of being defined, although perhaps not specifically by the power of the bundle *a partir de sus*.

Describing

Example 3 shows a footnote which is one of the few examples of a Describing sentence in the dataset.

- (3) For useful discussions of the complexities of settler colonization **in the United States** and Israel, respectively, see Dolores Janiewski, “Gendering, Racializing, and Classifying: Settler Colonization in the United States, 1590–1990”; and Nahla Abdo and Nira Yuval-Davis, “Palestine, Israel, and the Zionist Settler Project,” in *Unsettling Settler Societies*, ed. Stasiulis and Yuval-Davis, 132–60 and 291–322. (375.txt)

Example 4 contains an example of Describing in the Spanish data.

- (4) El relato explora el proceso de las independencias latinoamericanas **en el siglo XIX** a través del personaje de Simón Rodríguez, preceptor de Simón Bolívar, convertido en Samuel Robinson por 26 años. (78.txt)
[The story explores the process of Latin American independence **in the 19th century** through the character of Simón Rodríguez (mentor to Simón Bolívar) who became Samuel Robinson for 26 years.]

The sentence contains two influential Latin American historical figures: Simón Rodríguez and Simón Bolívar. Readers familiar with the lives of both these men would not necessarily need information on the century in which they lived, but the remainder of readers benefit from this bit of contextual elaboration. Here, then, we see possible evidence at the sentence level for increased prepositional phrases per million words occurring at higher rates as rates of academic elaboration, measured by dimensional scores in past studies in Spanish (Biber et al., 2006), increase.

Retelling

Example 5 was “Retelling”, perhaps due to the high quantity of verbs *consolar* [to console], *llorar* [to cry], *augurar* [to foresee], and *ver* [to see], the final three of which occur in a relative clause triggered by the lexical bundle itself.

- (5) no ha podido consolar al Ángel **de la historia que** lloraba por un mundo roto por su eje y les augura algo mejor que lo que él ha visto: «Tampoco habéis visto lo enrojada de sangre que está esa esfera en algunas partes. Ni Dios quiera que la

veáis nunca, y no podáis ver partido el eje del mundo, ni al Ángel desconsolado» (2013: 11). (156.txt)

[He has not been able to console the Angel **of the story who** was crying for a world broken by its axis and he predicts something better than what he has seen: «You have not seen how reddened that sphere is in some parts either. God forbid that you ever see it, and you cannot see the axis of the world broken, nor the disconsolate Angel » (2013: 11).]

With regards to the bundle used, the polysemy of *historia* in Spanish can refer to both history proper, as well as a story or an anecdote (Diccionario de la Real Academia Española). In this case, it appears to refer to a story and, co-occurring with a wide range of verbs, qualified it for retelling. It is possible that future studies could explore the idea that bundles that cluster predominantly to Retelling functions may exist, but may not occur as often.

Explaining

Sentences in the dataset containing LBs which use the trigram *a través de* tended to work toward Explaining, as shown in Examples 6-7.

- (6) En realidad, el observador inglés es una trasposición literaria del Baroja exiliado en París, que recibe noticias de Madrid **a través de las** visitas de los amigos a la capital francesa, y los relatos le permiten hacer una serie de reflexiones sociales e históricas. (156.txt)
[In reality, the English observer is a literary transposition of Baroja exiled in Paris, who receives news from Madrid **through** visits by friends to the French capital, and the stories allow him to make a series of social and historical reflections.]

Example 6 explains details that would not be obvious immediately to a reader without using subjective language.

Example 7, also rated as “Explaining”, elaborates on a novelist’s techniques.

- (7) Jiménez Lozano, con otra modalidad literaria afín a la epistolar, el acercamiento a las voces de los personajes **a través de una** mediación amorosa y fiable —la del sobrino—, finge reproducir un diálogo realizado oralmente. (156.txt)

[Jiménez Lozano, with another literary modality related to the epistolary, the approach to the voices of the characters **through** a loving and reliable mediation - that of the nephew -, pretends to reproduce a dialogue carried out orally.]

Here, the bundle *a través de una* performs a peripheral referential function attached to the noun phrase “el acercamiento a las voces de los personajes” [the approach to the characters’ voices].

Example 8 is an example of an Explaining token in English.

- (8) The withdrawal of life from the obligations of the community paradoxically engenders a transcendental sovereignty which has the power to sacrifice and extinguish life **in the name of** the preservation of life. (45.txt)

While there may not be many inherently lexicogrammatically or semantically explanatory features in the bundles shown in this section, remarks on the prepositional phrases involved are a topic to which we return at the end of this section.

Evaluating

Examples 9-11 appear to be examples of Evaluating from the English-language data.

- (9) Abu-Jaber’s portrayal of white working-class life thus **can be read as** a critique of U. S. society, a refutation of a national imaginary that links the United States with progress, endless possibility, and the temporality of the future. (375.txt)
- (10) Similarly, Sapphire’s previous claim that women must push back against violence **can be read as** a call for women to rally against gender-based violence. (471.txt)
- (11) Johnson’s Europe **can be read as** a combination of an early New Negro dedicated to the advancement of the race and a dedicated and innovative New Negro artist. (314.txt)

With regards to Evaluating sentences containing stance bundles as seen in Examples 9-11, the Spanish *el hecho de que* [the fact that] stands out in Examples 12-13.

- (12) A todos estos datos empíricos se sumaría **el hecho de que**, bajo mi punto de vista, Andrés Hurtado es también un alter ego de Baroja. (155.txt)
[To all these empirical data would be added **the fact that**, from my point of view, Andrés Hurtado is also an alter ego of Baroja,]

The example *bajo mi punto de vista* [from my point of view], (which, despite not occurring frequently enough to qualify as an LB, could still be plausibly argued as a stance marker) appears to be the most salient feature of Evaluating according to the criteria in Appendices A & B.

Example 13, while containing the bundle *el hecho de que* [“the fact that”], appears to be Evaluating due to the presence of the overt marker of the author’s opinion “me parece absurdo” (it seems absurd to me)”.

- (13) En este sentido, y aceptando que estas afirmaciones se hacen en textos —estos sí— totalmente autobiográficos y no ficcionales, leídos en función de un «pacto autobiográfico» entre autor y lector, me parece absurdo negar la evidencia de que, efectivamente, la razón más poderosa para hablar de *El árbol de la ciencia* como de una novela autobiográfica reside en **el hecho de que** su protagonista no es otra cosa que la encarnación literaria, la versión en la ficción, de la persona y la personalidad del autor durante una fase de su vida. (155.txt)
 [In this sense, and accepting that these affirmations are made in texts - indeed - totally autobiographical and non-fictional, read based on an "autobiographical pact" between author and reader, it seems absurd for me to deny the evidence that, in effect, the most powerful reason to discuss *The Tree of Knowledge* as an autobiographical novel lies in **the fact that** its protagonist is none other than the literary incarnation, the version in fiction, of the person and personality of the author during a phase of his life.]

As Table 3.21 reported, in Section 3.6.1 of Chapter 3, the average number of words per article qualifying as a four-word bundle in Spanish was 3.6% (SD = 1.2%) and 1.4% in English (SD = 0.6%). Given the minor presence of four-word LBs compared to the remainder of the prose in a given articles, it is possible to consider Examples 9-13 as lending support hypothesis that some bundles, such as *can be read as a* may instantiate a communicative purpose. Other bundles, such as *el hecho de que*, or indeed, its counterpart *the fact that the*, while not the most salient marker of stance in a sentence, may co-occur with other stance markers in informing the larger communicative purpose. It has been found that the majority of lexical bundles in academic writing are noun phrases in conjunction with a preposition (Biber et al., 1999). As such, it is

possible that some lexical bundles, such as those in the minority that contain verb phrases, will hold more communicative salience than others.

Comparing

Example 14, a Comparing sentence, earns its status based on the phrase “La diferencia entre don Quijote y Lodaes”.

- (14) La diferencia entre don Quijote y Lodaes es su conciencia del mundo: la melancolía de Lodaes nace del desengaño ante un mundo roto y ensangrentado, mientras que don Quijote pone de manifies to algunas injusticias de un mundo humano **a través de sus** descabelladas aventuras. (156.txt)
 [The difference between Don Quixote and Lodaes is their awareness of the world: Lodaes’ melancholy arises from disappointment in a broken and bloody world, while Don Quixote reveals some injustices of a human world **through** his hare-brained shenanigans.]

It appears that here, the bundle *a través de sus*, while not connected to the establishment of the difference between the two characters, occurs when providing contextual information about don Quijote. Bundles such as this one, containing even more prepositions, may, in some indirect fashion, gravitate toward scholarly inferences thanks in part to the association between greater preposition use and dimension scores more characteristic of academic writing in English and Spanish (Biber, 1988; Biber et al., 2006).

Example 17 shows an English-language instance of Comparing:

- (17) The reference to maniacs throwing acid provides an intertextual link to Himes’s Coffin Ed, whose face is scarred **as a result of** a freak’s throwing a bottle of acid at him, an event that is recorded in the Serie Noire novel titled Lareine des pommes. (318.txt)

Although the bundle occurs when summarizing plot information, the Comparison lies in the fact that the plot information is used to support the idea of an intertextual comparison. As discussed in Chapter 4, higher levels of prepositions have been found to be more strongly associated with

informational, rather than conversational, language in English (Biber, 1988) and Spanish (Biber et al., 2006). We entertain here the possibility that, given Intangible Framing Attributes' tendency to contain at least one preposition (Tracy-Ventura et al., 2007), while they may not necessarily convey more scholarly discussions of information, their occurrence may be correlated with a literary scholar's doing so.

6.3 Equivalent Bundle Communicative Purpose Analysis

Concerning equivalent LBs in meaning and function across languages, the following examples are shown in Table 6.3.

Table 6.3. Equivalencies between bundle and communicative purpose across languages.

Bundle pair	Comm. Purpose	Spanish	English
desde la perspectiva del & from the perspective of a	Explaining	La creación de dos narradores observadores permite a Baroja guardar una distancia respecto a lo narrado y dar la versión de los acontecimientos desde la perspectiva del extraño —del extranjero, del que viene de fuera— y, como tal, tiene libertad para mezclarse y tratar de comprender las dos facciones. (156.txt)	He is bare life, the man without rights from the perspective of a sovereign power protecting democracy; but he is also the absolutism of the sovereign individual, radically dispensed from the obligations of community. (45.txt)

Table 6.3 (Continued)

en el caso de & in the case of	Explaining	<p>Aunque en el caso de algunos de ellos esta relación no quede del todo clara, en otros resulta evidente el paralelismo y no es difícil buscar en las memorias o en los escritos autobiográficos de Baroja el posible modelo de cada uno de estos tipos literarios. (155.txt)</p> <p>[Although in the case of some of them this relationship is not entirely clear, in others the parallelism is evident and it is not difficult to search in Baroja's memoirs or autobiographical writings for the possible model of each of these literary types.]</p>	<p>In the case of the United States, these factors give rise to an idealized notion of U.S. citizenship that is grounded in the logic that we all suffer in the same way and thus must deploy similar strategies to overcome hardship. (471.txt)</p>
de la novela es & of the novel is	Evaluating	<p>Si entendemos al personaje autobiográfico en el sentido más ortodoxo, como un trasunto de la persona real del autor que revive las situaciones anteriormente vividas por este y adopta en la novela sus actitudes y opiniones, creo sin duda alguna que el único personaje autobiográfico de la novela es el protagonista. (155.txt)</p> <p>[If we understand the autobiographical character in the most orthodox sense, as a transcript of the author's real person who relives the situations previously lived by him and adopts his attitudes and opinions in the novel, I believe without a doubt that the only autobiographical character in the novel is the main character.]</p>	<p>These diverse approaches to interpreting the novel – as ethnography, coming-of-age narrative, and/or morality tale – seem to be contradictory; the culture and historical context of the novel is the object of the narrative or merely the backdrop for a universal story. (562.txt)</p>
el hecho de que & the fact that the	Evaluating	<p>En este sentido, y aceptando que estas afirmaciones se hacen en textos —estos sí— totalmente autobiográficos y no ficcionales, leídos en función de un «pacto autobiográfico» entre autor y lector, me parece absurdo negar la evidencia de que, efectivamente, la razón más poderosa para hablar de El árbol de la ciencia como de una novela autobiográfica reside en el hecho de que su protagonista no es otra cosa que la encarnación literaria, la versión en la ficción, de la persona y la personalidad del autor durante una fase de su vida. (155.txt)</p>	<p>The fact that the victim's throat is cut, or "snicked," by the word, from the inside, is at once comic and sad. (318.txt)</p>

Table 6.3 (Continued)

[In this sense, and taking at face value that these statements are made in totally autobiographical, non-fictional texts —yes, texts—and, read based on an “autobiographical pact” between author and reader, it seems absurd to deny the evidence that, indeed, the most powerful reason to speak of *The Tree of Knowledge* as an autobiographical novel lies in the fact that its protagonist is nothing other than the literary incarnation, the fictional version, of the person and personality of the author during a phase of his life.]

Near equivalencies between bundle and communicative purpose appear in Table 6.4.

Table 6.4. Partial agreement in communicative purpose judgments.

Bundle	Sentence	Comm. Purpose	Sentence	Comm. Purpose
en relación con la & in relation to the	Toda la trayectoria de Lodaes, su existencia concreta en el siglo XX, está vivida en relación con las Cruzadas, y lo que resulta de esta comparación es un juicio durísimo sobre el momento de estas nuevas «cruzadas» que le ha tocado vivir, las de la Guerra Civil española desde los dos bandos. (156.txt) [The entire history of Lodaes, its concrete existence in the twentieth century, is lived in relation to the Crusades, and what results from this comparison is a harsh judgment on the timing of these new “crusades” that he has had to live, those of the Spanish Civil War from both sides.]	Evaluating	In this article, I do not provide a critical reading of the novel as a window into Afghan culture or in relation to the long tradition of western representations of Afghanistan, as described by Fowler, for instance. (562.txt)	Explaining

Table 6.4. (Continued)

la forma en que & the way in which	<p>Me refiero al personaje de Lulú, de quien se ha dicho que «tiene de don Pío su sentido moral de la lealtad por encima de todo» (Medrano, 2006: 325), y que su mayor interés con vistas a descubrir el trasfondo autobiográfico de la novela son «sus cualidades morales e intelectuales, y la forma en que Baroja expresa, [...] (155.txt)</p> <p>[I am referring to the character of Lulú, who has been said to have “Don Pío’s moral sense of loyalty above all else” (Medrano, 2006: 325), and that his greatest interest with a view to discovering the autobiographical background of the novel is “his moral and intellectual qualities, and the way that in which Barjoa expresses [...]]</p>	Explaining	<p>I critically examine Amir’s need “to be good again” in order to analyse the way in which (the desire for) a global-humanitarian ethics underwrites the historical context of The Kite Runner’s production and popularity. (562.txt)</p>	Evaluating
--	--	------------	---	------------

Increased prepositional phrases were shown in both Spanish (Biber et al. 2006) and English (Biber, 1988) MDA studies to be more characteristic of academic language. While the majority of LBs discussed in this section begin with prepositional phrases, the equivalent bundle pairs that buck the trend are *el hecho de que* & *the fact that the* and *la forma en que* & *the way in which*. Increased relative clauses and noun complement clauses in both MDA studies were found to be more characteristic of interviews than of academic texts. Taken together, the equivalent analysis offers cross-linguistic support that prepositional phrases, noun-complement clauses, and relative clauses instantiated with lexical bundles may be one viable strategy of enacting more inferential communicative purposes.

6.4 Agreement results

As discussed in Section 3.4 of Chapter 3, a study to determine the extent to which LBs can be perceived as linked inherently to communicative purposes was conducted with four participants. Two participants were doctoral students in an English literature program and two in a Spanish literature program. The specific training materials used to orient the participants are available in Appendices A-E. In accordance with the agreement measures defined in Section 3.4.3 of Chapter 3, agreement was calculated by the average agreement score of all tokens containing the same bundle using the following formula summarized in Table 6.5.

Table 6.5. Agreement calculation formula

Category	Conditions	Points
Full	All participants agree with the researcher	1.00
3/4	The researcher agreed with at least one of the participants	0.75
Half	The researcher agreed with both raters that the sentence was somewhere on the “reporting” or “inferring” end of the hierarchy	0.50
1/4	The researcher agreed with one rater that the sentence was somewhere on the “reporting” or “inferring” end of the hierarchy	0.25
None.	No exact or adjacent agreement between the researcher and the other two raters	0

Thus, for example, the single token of *as a form of* in Table 6.6 reveals that the researcher agreed with only one rater that the sentence was somewhere on the “reporting” or “inferring” end of the hierarchy. The token of *a partir de sus* [from hers/his/theirs/its] in Table 6.6, however, with an agreement of 100%, shows that all raters agreed.

Table 6.6 shows the agreement scores generated for the sentences in which each English-language bundle occurred.

Table 6.6. Mean agreement per English bundle.

Bundle	Tokens	Agreement	Bundle	Tokens	Agreement
as a form of	1	25%	in the form of	2	75%
as a result of	1	75%	in the name of	1	50%
as one of the	1	50%	in the United States	4	62.5%
as part of a	1	25%	it is important to	1	25%
as part of the	1	0%	of the novel is	2	50%
as well as a	1	50%	on the level of the	2	12.5%
as well as the	1	75%	on the one hand	3	50%
at the heart of	2	62.5%	on the other hand	3	75%
at the same time	3	33.3%	that there is a	1	75%
can be read as	3	83.3%	the end of the	1	75%
from the perspective of					
a	1	0%	the fact that the	2	0%
in a way that	1	0%	the figure of the	3	75%
in relation to the	2	75%	the role of the	2	37.5%
in the case of	2	50%	the way in which	3	66.7%
in the context of	1	0%	Mean		51.0%
			SD		26.1%

Spanish results of agreement by bundle are shown in Table 6.7.

Table 6.7. Mean agreement per Spanish LB

Bundle	Tokens	Agreement	Bundle	Tokens	Agreement
a la vez que	1	50%	en el que se	2	38%
a partir de la	3	75%	en el sentido de	1	25%
a partir de sus	1	100%	en el siglo de las		
a partir de una	1	50%	lucis	1	0%
a pesar de la	1	50%	en el siglo xix	1	100%
a pesar de que	1	50%	en la medida en que	1	25%
a pesar de su	2	25%	en la novela de	1	0%
a sí mismo como	1	50%	en la novela la	3	67%
a través de las	4	94%	en la novela que	2	50%
			en la obra de	3	58%

Table 6.7 (Continued)

al final de la	1	75%	en la que se	1	75%
cada uno de los	1	25%	en la vida de	1	75%
de la casa de	1	25%	en las que se	2	50%
de la historia de	2	50%	en lo que se refiere a	1	25%
de la historia que	1	100%	en los que se	1	75%
de la historia y	2	75%	en relación con la	1	0%
de la muerte de	1	25%	en torno a las	1	75%
de la novela como	1	50%	la construcción de un	1	25%
de la novela en	2	64%	la forma en que	1	75%
de la novela es	1	50%	la historia de la	1	100%
de la novela que	2	25%	la historia y la	2	50%
de la novela y	1	0%	la lectura de la novela y	1	75%
de la obra de	2	50%	la mayoría de los	1	25%
de la vida de	2	63%	la posibilidad de un	2	50%
de las novelas de	1	75%	la problemática de la	1	50%
de los personajes que	1	75%	la vida y la	1	50%
de una serie de	1	0%	más allá de la	1	50%
desde el punto de vista	1	50%	no deja de ser	1	0%
desde la perspectiva del	1	100%	parte de la novela	1	0%
el hecho de que	2	100%	parte de la novela	1	38%
el momento en que	4	44%	que se refiere a	1	25%
el protagonista de la novela	4	38%	se convierte en un	1	75%
el punto de vista	1	25%	se manifiesta en las	1	75%
en el caso de	9	53%	se trata de la	1	25%
en el momento de	1	75%	se trata de un	1	75%
			y por lo tanto	1	50%
			Mean		53%
			SD		25%

Both Tables 6.6 and 6.7 reveal agreement scores of only slightly over 50%. Remarks on why agreement scores varied are a topic to which we return in Section 6.6.

6.5 Saliency Results

We defined Saliency of lexical bundles as follows: Each of the three raters who mentioned a bundle in their response by name contributed a score of +0.33 to the bundle saliency score for each sentence in the dataset. This process is summarized with the following formula:

$$\begin{aligned} & ((\text{Researchers who mentioned the bundle in Rationale column} * 0.33) + \\ & (\text{Researchers alluded to the content of a bundle} * 0.165)) = \\ & \underline{\text{Bundle saliency score.}} \end{aligned}$$

In line with Research Question #4 which asked whether some LBs cluster to certain communicative purposes, it was hypothesized that some bundles may be salient enough to readers and set the stage for higher agreement scores. However, after looking at agreement patterns in the agreement analysis, a Spearman's Rho found no significant correlation between bundle salience and agreement. However, it was hypothesized that agreement and Saliency might vary according to functional category. To that end, Table 6.8 shows means and SDs for agreement and bundle salience by functional category, based on categories used in past English- and Spanish-language bundle studies (Biber, Conrad & Cortes, 2003; Tracy-Ventura et al., 2007; Cortes, 2008).

Table 6.8. Mean agreement and salience by functional bundle category in English.

Category	N	Agreement	SD	Saliency	SD
Intangible Framing Attributes	22	48%	34%	12%	14%
Multi-Function	2	38%	53%	0%	0%
Obligation / Directive	1	25%	0%	67%	0%
Place-Institution	4	63%	43%	0%	0%
Stance Epistemic Impersonal	8	53%	39%	46%	30%
Tangible Framing Attributes	2	75%	35%	0%	0%
Time-Event	3	33%	29%	61%	19%
Table 6.8 (Continued)					
Topic Elaboration / Clarification	10	58%	21%	13%	17%

Table 6.9 does so for the Spanish bundles by category.

Table 6.9. Mean agreement and salience by functional bundle category in Spanish.

Category	N	Agreement	SD	Saliency	SD
Epistemic Impersonal	4	83%	29%	4%	8%
Identification / Focus	7	47%	31%	5%	13%
Intangible Framing Attributes	42	52%	34%	17%	21%
Multifunctional Reference	9	83%	18%	7%	17%
Place-Institution	1	100%	0%	17%	0%
Quantity Specification	1	25%	0%	33%	0%
Referential Tangible	1	0%	0%	50%	0%
Structural Only	6	54%	19%	14%	27%
Tangible Framing Attributes	23	47%	30%	21%	2%
Time-Event	7	54%	27%	17%	37%
Topic Elaboration-Clarification	4	31%	24%	13%	25%

The results reveal that the primary functional category with any degree of correspondence between agreement and salience comes in the form of English-language Stance Epistemic Impersonal bundles.

6.6 The tactile experiencers and the elephant: why raters sometimes disagreed

Tables 6.8 and 6.9 show wide variation in interrater reliability for the dataset analyzed in this chapter. The present section represents an attempt to explore potential problematic areas when exploring rater reliability measures in this type of analysis. Reviewing academic literature on humor, Berger (2011) used a South Asian parable as a metaphor to describe several academic disciplines' distinct approaches toward explaining the reasons for which humans experience delight. This parable tells of several men who live with visual impairments. Each man, with a hand touching a separate part of a live elephant, formulates a distinct opinion as to the identity of the elephant based on the portion of the elephant's body that each one can touch. For example, the man holding on to the elephant's trunk has a difficult time convincing the others that they are dealing with a snake.

A similar phenomenon appears to have manifest between each set of participants and the researcher. That is, all raters disagreed on the communicative purpose of Example 18, but perhaps for worthwhile reasons.

- (18) **On the one hand**, Professor's decision to join the 369th epitomizes what Chad L. Williams terms African American troops' desire to demonstrate "the broader manhood of the race through the heroic valor of military service" (7). (314.txt)

Participant A-EN (Alias "Kay") selected Defining, deeming the most salient portion as "what [...] terms", treating the subsequent phenomenological construct as a term. The researcher, in contrast, wrote the following rationale for Comparing, with information on the lexical bundle occurring to influence the perception: "Comparing actions to remarks of someone else and using it as an example. So it could be explaining too." Participant B-EN (alias "Robin"), in turn, wrote a rationale suggestive of Comparing, but chose Explaining, describing the sentence as "Making

connections between the novel's events and the novelist's themes." Thus, there is a case to be made that each of the raters made principled inferences on the communicative purpose of Example 18, but based on distinct segments of a 36-word sentence. It appears then that when raters disagree, one reason is due to seeing distinct segments of the same sentence as most salient, just like the ancient South Asian men describing distinct parts of an elephant to each other based on their senses of touch. In the following section, we examine the characteristics of high rater agreement.

With regards to targetlike agreement levels, Krippendorff (1980) writes that raters should: "[c]onsider variables with reliabilities between $\alpha = .667$ and $\alpha = .800$ only for drawing tentative conclusions" (p. 241). Keeping these statistics in mind, we also note agreement for some bundles, i.e., *can be read as* and *el hecho de que* were among the highest. Another cause for discrepancy in ratings may stem from the fact that the participants had degrees in literature and were thus less familiar with approaching disciplinary texts from linguistic viewpoints.

6.7 Conclusion

In summary, this chapter has examined to what extent lexical bundles could be associated with specific communicative purposes. Specifically, it has been found that there is a larger attestation for the Communicative Purposes Hierarchy used by Lake & Cortes (2019) across the corpus used in the present study. Further, there is initial evidence to suggestion that LBs in the Stance Epistemic category cluster toward inferential communicative purposes in both languages, with support from the trained raters' evaluations. some lexical bundles, such as those in the

Stance Epistemic category stand out particularly for their levels of agreement and for their cross-linguistic equivalency.

In determining parameters for an enhanced approach toward future LB communicative purpose analyses, some related remarks for context over raw lexicogrammatical features are in order. Consider, for example, the Swalesian move “Contextualizing Study Methods” that occurs within a methods section (Cotos et al., 2017). Lake & Cortes (in press) found that the LB *de la ciudad de* [from the city of] occurred when researchers contextualized their own study methods. However, the possibility remains that another researcher contextualizing someone else’s study methods would, by convention, do so when summarizing previous research (Jian, 2010). A similar argument could be made for lexical bundles occurring in Results sections. There may exist bundles that could be used by a researcher to describe her or his own findings in a Results section as well as by another researcher to summarize them in a literature review afterwards. The register analysis in Section 4.1 of Chapter 4 revealed that, while few articles in LCW explicitly mark introductions and conclusions, most do mark section headings. There may exist at least a possibility that some lexical bundles could occur in Introductions and/or the first explicitly marked section in a given research article, as well as in the final, explicitly marked section. For example, Lake & Cortes (in press) found that the LB seemingly devoid of any inherent specificity *cada uno de los* [each of the] occurred significantly more often when describing methodological procedures in Spanish-language psychology articles than in any other section. Thus, the distribution of similar bundles between abstracts, introductions, conclusions, and other potentially marked sections in LCAs warrants further testing in future studies.

Additionally, future studies could present participants with data that has the lexical bundle in each sentence appearing in bold and asking participants to focus on what is achieved by the bundle in a given sentence. In this way, returning to the parable of the blind men and the elephant, participants who, figuratively, disagreed on differences in the ivory in the tusks or the keratin in its eyelashes could be moved to all examine the trunk, or in this case, the communicative purpose of a lexical bundle, simultaneously and in harmony.

7 CONCLUSION

7.1 Overview

Corpus-driven (Biber, 1988; Biber, Conrad & Cortes, 2003; Parodi, 2005; Gray, 2015) and mixed-methods analytical frameworks (Swales, 1990; Swales & Feak, 2004) have enhanced the teachability of academic writing in a variety of disciplines. Although some syntactic, corpus-driven (Biber, 1988), and mixed-methods (Tankó, 2017) studies have been performed on literary criticism, none of them, to the author's knowledge, apply lexical bundles, MDAs, and mixed methods approaches specifically to literary criticism. Addressing this need, this dissertation analyzed two corpora of expert-level LCW in English and Spanish. The articles therein analyzed a single novel. The results of the MDA, discussed in Chapter 4, reveal that LCW at times resembles several varieties of academic prose, while also taking on features of religious texts, broadcasts, and official documents in English, in addition to resembling public political speeches in Spanish at times. Chapter 5 demonstrated the existence of both cross-linguistic shared and language-specific utterances that contribute to the production of an LCA which has undergone peer-review. Chapter 6 showed some support for the mapping of communicative purposes to lexical bundles, while finding also that there are ways to improve findings in this regard in the future. These findings make novel contributions to extant knowledge on the kinds of linguistic features involved in an expert-level LCA.

In this chapter, the dissertation is concluded with a review of the study's key findings (Section 7.2), an assessment of the study's limitations (Section 7.3), its pedagogical implications (Section 7.4), and future directions for research (Section 7.5).

7.2 Summary of Findings

7.2.1 MDA-based register classification of literary criticism in English and Spanish

The MDA results generated from the English and Spanish corpora provide a focused description of the linguistic features characteristic of LCW by placing their relative levels of occurrence in relation to other registers. To illustrate, while English LCAs had many of the hallmarks of research articles, such as high levels of nominalizations, LCW parts ways from other types of academic prose most notably in its subtle, barely observable persuasive devices. There is reason to continue investigating the extent to which passive voice constructions are used for a phenomenon we might call “covert expression of persuasion”. LCW also more closely resembles a religious text than social science academic or humanities academic prose, due perhaps to the quantity of integrated citations (Hyland, 1999). Aside from those differences, English LCAs resemble research articles in other disciplines in their preference for high levels of preposition- and noun-phrases, the present tense, and a moderate degree of passive constructions.

Similarly, while Spanish LCAs exhibited many of the previously attested norms in Spanish-language academic writing, they stood out saliently for their heightened levels of interactive, contextual adverbs, which are more characteristic of political speeches than of research articles, perhaps as a consequence of relying more on argumentation than on tables to present evidence.

Spanish LCAs are slightly more narrative than other research articles, due to their opting occasionally for the imperfect tense when describing plot details. LCAs in Spanish tend to prefer the historical present, leaving them distinct from the narrative focus levels of the novels which they seek to analyze. They are also less committal than other documents, like research articles in

other fields, opting for fewer third-person pronouns or use of the preterite tense. LCAs' levels of hedging verbs, adverbs of manner, and predicative adjectives also cause them to resemble political speeches and works of literature more than they do academic research articles in this regard, cementing their status as a register which takes on both literary and scientific features.

7.2.2 Function and structure of LBs in literary criticism

The analysis revealed that the Spanish corpus contained more than double the amount of unique LB types attested in the English corpus. Additionally, 60% of these LBs in Spanish and 40% in English may be LCW-specific. Further, similar to findings in past English (Biber et al., 1999) and Spanish (Cortes, 2008) LB studies, prepositional phrase bundles prevail in both languages. Spanish, compared to English, has higher quantities of that-clause bundles, verb phrase, and adverbial phrase bundles. English, on the other hand, has one question word bundle, but comparable levels of prepositional phrase bundles. Functionally, Spanish had a greater number of Intangible framing attributes bundles, Tangible framing attributes, Structural, and Multifunctional bundles. English, however, predominated with Topic elaboration/ clarification, Stance, Time-event, Obligation/directive, and Deixis bundles.

Similarities were found in the kinds of lexical bundles used in both languages in the form of Time-Event bundles. Quasi-similarities were found in the utilization of stance bundles, although these particular findings opened up further discussions about cross-linguistic differences observed in the field of applied linguistics (Perales Escudero & Swales, 2011) that also appear to be at play in LCW. Language-specific bundles offered a point of departure for discussions on advanced prepositional constructions in Spanish.

7.2.3 *Communicative purposes of lexical bundles chosen*

The pilot analysis revealed that LCW contains both descriptive and inferential communicative purposes. The overwhelming majority of bundles occurring in the sample clustered toward inferential communicative purposes. In particular, strong support was found for the idea of stance bundles in both languages playing a noticeable part in the formulation of evaluations in the bundles *can be read as*, as well as *el hecho de que* [the fact that].

Additionally, LBs were found to occur in sentences which define key terms and draw comparisons. However, it appears at the sentence level that the LBs' roles in these sentences are more peripheral and contextual, rather than inherently linked to said communicative purposes. Further analyses of LB patterns at the paragraph level may reveal more observable Defining and Comparing phraseological sequences facilitated directly by LBs in both languages.

Lastly, referential bundles linked to times and places (i.e., *in the United States*, *en el Siglo XIX* [in the 19th century]) were found to occur in the context of Describing. Increased levels of prepositional phrases are more characteristic of academic texts than of novels in both languages (Biber, 1988; Biber et al., 2006). Thus, a selection of referential bundles may prove to be useful in further pedagogy on the production of objective, academic descriptions in LCW.

7.3 Limitations

This dissertation exhibits limitations which could be remedied by future studies. First, while MDAs were conducted in Spanish and English, the English-language analysis measured levels of 51 linguistic features, whereas the Spanish approach used only 23. One contributing

factor to this discrepancy is the fact that, whereas Venegas (2010) reported mean dimension scores that facilitated methodological replication, the Spanish-language MDA conducted by Biber et al., (2006) on a larger amount of linguistic features did not. More referential data on the norms of Spanish-language LCW could become available if another MDA of the current iteration of the *Corpus del español*, now composed of four times the amount of texts that it did at the time of Biber's et al. (2006) analysis, were conducted. With such data, proper significance testing of dimensional scores using ANOVAs could also be properly conducted to obtain a clearer picture of the extent to which LCW differs from other registers. Additionally, a new MDA generated entirely from intra-corpus distributions of linguistic features (cf. Asención-Delaney, 2014), rather than the approach taken in this dissertation which places findings put in dialogue with frequency of occurrence norms for linguistic features occurring in norms codified by past MDAs (i.e., Biber, 1988; Parodi, 2005), could provide further knowledge that would complement the description of LCW in both languages.

Second, while the lexical bundles were analyzed functionally and structurally, more contextual data on the stage at which they tend to be used is warranted. A worthwhile starting place would be to replicate Lake & Cortes (in press), who measured the variation in rates of occurrence of lexical bundles in Psychology article Abstract, Introduction, Methods, Results, and Discussion sections. Although LCAs do not use IMRD heading labels, the situational analysis in Section 3.1 of Chapter 3 revealed that over 2/3 of the articles in both corpora used some manner of section headings. Thus, if the corpora were segmented in such a way as to have the abstract of each text in a separate text file, first labeled section as the introduction as the introduction, the final labeled section as the conclusion, and all occurring between as the body, significance tests

could be conducted to obtain more insight on the kinds of rhetorical moves that occur in a publishable LCA.

Finally, the researcher and the raters in both languages tended to base their communicative purpose judgments on distinct segments of the same sentence. One reason that this discrepancy occurred was that no input enhancement (White et al., 1991) was performed on the bundles (i.e., bolding, underlining, or italicizing). Future studies could replicate the methods herein with the lexical bundles made more noticeable and asking the participants specifically, “what communicative purpose do you think that this phrase is helping to communicate in this sentence?” In that way, participants may become more inclined to base their judgments on the same portions of a given stimulus. Contextual information on semantic preferences and prosodies (cf. Cortes & Hardy, 2013) could also contribute tighter methodological parameters for instruments used in future LB communicative purpose studies.

7.4 Pedagogical Implications

Although LCW is a ubiquitous genre in the Spanish-language degree program (Hertel & Dings, 2014; 2017; Small, 2015) and the subject of a call for effective educators to become equally developed as humanities scholars (Francomano, 2012), there is a dearth of previous studies on LCW using the empirical approaches that have yielded data to enhance instructional materials for other academic disciplines.

This dissertation’s findings suggest that while LCW in English and Spanish resembles academic prose in many regards, it also contains many observable hallmarks that cement its status as distinct, requiring unique skills to produce. The analysis which responds to the first

element RQ1, described in Section 4.2 of Chapter 4, shows that LCW, like other academic disciplines, favors a noun-dense, present-tense, style which, contrary to past assessments of humanities writing (Biber, 1988), favors the passive voice more than other genres. Further, LCW elects to utilize, rather than tables, explicit prose involving arguments rich with nominalizations, adverbs and integrated citations using noun-, verb-, and adjective-complement clauses. However, curiously, LCW expressed persuasion less overtly than other kinds of writing, such as social science academic prose.

Together, these results suggest that, for English LCW, instructional materials could offer the following as complements to extant academic writing pedagogical materials. Materials designers could create tasks showing frequent noun (*the fact/idea/assertion/claim that*), adjective (*is clear/evident/significant/etc.*), and verb (*suggests/argues that*) collocates of complement clauses as recommendations for deploying integrated citations (Hyland, 1999) of primary and secondary source texts in an LCA. Further, materials designers now have justification to state that overt expressions of persuasion are seldom present in LCW. They could thus encourage learners to practice a more indirect presentation of evidence in support of their theses using passivization strategies, i.e., “Nobantu’s loss of touch with reality can **be interpreted** as being a punishment from the ancestors of the amaMpondomise, as she never respected them after marrying Zwelinzima.” (Kwatsha, 2007, p. 86). It is clear that the interpretation put forth is an idea which the author intends to express. However, it is presented with the modal *can* rather than *should*. Further, in such examples, materials designers could encourage learners to be maximally explicit with a combination of highly specific nominalizations which are qualified to specific degrees using adverbs, as in the case of “Jazz became **more** popular in France after World War II

partly because it had become a ‘cultural symbol of **antifascism**’ during the **Occupation** (Stovall 134).” (Lowney, 2015, p. 115). In this example, for instance, the author uses both specific words such as occupation and antifascism, in concert with a qualification of the statement using the adverb “partly”, acknowledging that the claim made in the sentence is not the sole reason for the popularity of Jazz in a postwar France, but nonetheless an important one. As such, instructors have further empirical support in encouraging students to qualify highly specific statements.

Similar remarks are in order for the second facet of Research Question #1 answered in Section 4.3 of Chapter 4. Similar to results obtained on Dimension 3 in English, the Dimension 1 results in Spanish show cross-linguistic support for the need to master the use of adverbs to qualify one’s statements, as an alternative to tables, to present evidence cited in support of the text’s thesis statement, as in the following qualified statement:

“Allí la grandeza del autor, a pesar que todo el libro habla sobre la muerte, termina con un canto a la esperanza, **no porque** re-signifique la muerte, sino **porque** espera que algún día los que piden pan **no** sean acusados de pecadores, asesinos o apátridas.”

[There, the greatness of the author, despite the whole book’s discussing death, finishes with a song of hope, not because it resignifies death, but rather because it hopes that some day those who ask for bread will not be accused of being sinners, murderers, or stateless persons] (Mansilla & Vélez, 2016, p. 147).

Similarly, like English, Spanish-language LCW uses the past tense to a degree much lesser than the novels which it analyzes, noting a preference for the present tense. The past imperfect may occur when describing historical context at times, however. The analysis also

revealed that Spanish LCW seldom utilizes such features as first-person declensions and preterite constructions, giving further support for encouraging a more impersonal style. Predicative adjectives, hedging verbs, and modal verbs of possibility do, however predominate, according to the Dimension 4 findings, suggesting that materials could teach favored predicative adjectives in the register (e.g., *evidente*, *importante*, *necesario*, *cierto*, etc. [evident, important, necessary, right) when qualifying one's statements. As for Dimension 5, these verbs were frequent in the subjunctive mood: *ver* [to see], *ser* [to be], *haber* [there to be], in conjunction with a wide array of nominalizations placed when putting forth claims. Encouraging student inquiry with corpora (cf. Bernardini, 2004) on guided tasks based on these findings could facilitate a sense of ownership of disciplinary conventions in both languages.

Concerning the first portion of Research Question 2, the findings suggest, when cross-referenced with past major studies on academic writing in English and Spanish (Biber, Conrad & Cortes, 2003; Tracy-Ventura, Cortes & Biber, 2007; Cortes 2008; Pérez-Llantada, 2014; Lake & Cortes, 2020), 60% of the LBs in English and 40% of those in Spanish revealed in this study appear to be specific to literary criticism as a discipline. Compared to proportions known for history writing (Cortes, 2008), LCW in both languages possessed fewer NP bundles, opting instead for more PP bundles, as well as visibly more VP bundles in both languages bundles in Spanish. Like history writing, Spanish and English contained a high degree of time-event bundles used to situate narrations of actions at a particular time. However, LCW in both languages contained more stance bundle types than did history writing (Cortes, 2008).

These results on lexical bundle use in LCW allow for a placement of the following pedagogical suggestions. Notable stance bundles in English contain features that appear to

contribute to the high Dimension 6 scores revealed in Research question 1, that is, noun-complement clauses (*by the fact that, despite the fact that, the fact that the, to the fact that*), verb-complement clauses (*it is worth noting that, this is not to say that*), as well as one bundle that can occur in adjective, noun, and verb-complement clauses *that there is a*. Exercises using these bundles to perform integrated citations (Hyland, 1999) would represent a worthwhile exercise in introducing learners to the disciplinary practices to demonstrate mastery before then being able to move on with using a thesaurus to allow more uniqueness in expression.

Similarly, although Spanish MDA did not examine complement clauses, the Spanish-language stance bundles reveal the possibility of similarly designed materials for Spanish. The stance bundle category in Spanish also included triggers for noun- (*el hecho de que, el hecho de que la/el/las/los, en el hecho de que el/la/los* [the fact that the]), adjective- (*si bien es cierto que el* [even though it is true that]), and verb-complement (*se da cuenta de que, no cabe duda de que* [he/she/they/it realize[s] that, there is not doubt that]) clauses. A comparative study of the rate at which complement clauses in both languages may aid in determining the utility of teaching these bundles is warranted. Nonetheless, that bundles performing these lexicogrammatical functions crossed the bundle frequency threshold in both languages indicates tentative support for their importance in disciplinary writing in both languages.

Concerning the second portion of Research Question 2, there is evidence for equivalent and quasi-equivalent bundles. Notable tendencies include the equivalencies in Time-Event bundles across both languages, indicating that situating a novel's events at a given time and place is a shared cross-linguistic similarity of LCW in Spanish and English. The bundles *al final de la novela* [at the end of the novel], *by/at the end of the novel, en la primera parte de* [in the first

part of], and *at the beginning of the* could be presented in building skills of resuming plot details in the present tense in both languages. Research question 2, in its exploration of stance bundles, also reveals interesting pedagogical exercises in equivalent, and yet language-specific ways of expressing stance as was seen in the equivalency pair of *no es más que* [is nothing more than] and *can be read/seen as*. That the former expresses more epistemic commitment than the latter can serve as a point of departure for calling to learners' attention patterns of linguacultural difference in LCW in both languages.

Finally, in reference to the third portion of RQ2, the affective intangible framing attributes bundles in English *in the wake/aftermath/face of* show a language-specific means of expressing attitude toward a given event being foregrounded in the past of the sentence in which these bundles are used. Meanwhile, the prepositional constructions involved in the Spanish-specific structural only, (in)tangible framing attributes, and time-event bundles offer a meaningful focal point at which to foreground complex prepositional constructions favored in Spanish-language academic writing.

In reference to Research Question #3, explored in Section 6.1 of Chapter 6, bundles that raters rated as being used to define, describe, or retell were few – only *in the United States, as well as a* in English clustered toward more descriptive communicative purposes. In Spanish, nonspecific prepositional phrase bundles such as *a partir de sus* [from hers/his/theirs/its], *a partir de una* [from a(n)] and *en el caso del* [in the case of the] were rated as performing more descriptive rather than inferential functions. Additionally, bundles containing the noun *historia*, *i.e., la historia de la, de la historia y, de la historia que*, were used for descriptive purposes, but also notably, the time-event bundle *en el siglo XIX*. While these findings warrant further testing,

there is evidence to suggest bundles dealing with the history or context of a given event and Time-event bundles, such as *en el siglo XIX* could be teachable for describing and defining concepts, as well as retelling historical or fictional events. Concerning the inferential range of bundles and communicative purposes, there is evidence that the bundle *can be read as* plays a central role in the realization of the communicative purpose evaluating. The remainder of the English-language bundles clustered toward inferential communicative purposes. However, their role may in the realization of the communicative purpose may be more instrumental. As for Spanish-language bundles assessed in Chapter 6, *desde la perspectiva del* [from the perspective of the] clustered toward explaining. Its role in outlining larger details about a given perspective could offer instruction on expanding the beneath-the-superficial-level inferences about a given concept.

El hecho de que [the fact that the] clustered toward evaluating, while *de la novela es* [of the novel is] could be used for both. In all, the results suggest that bundles may either instantiate directly or occur indirectly as a correlate of a given communicative purpose. As for Research Question #4, as the heat maps in Appendices P and Q shows, while a full range of communicative purposes could be identified in the dataset, the communicative purposes of the majority of sentences in English and Spanish were classified as Explaining and Evaluating. The empirical study described in Sections 3.4 of Chapter 3 and Section 6.4 of Chapter 6, in which literary critics were asked to categorize the communicative purposes of sentences containing lexical bundles, should be replicated on a larger scale testing a larger number of the bundles. Nonetheless, the results suggest that a measurable inventory of lexical bundles constitutes the toolkit for instantiating scholarly explanations, evaluations, and comparisons. Further, these

findings could be used to design materials that scaffold the process of including higher-order argumentation in a literary criticism assignment.

7.5 Suggestions for Future Research

Although this dissertation has contributed to extant knowledge on literary criticism writing, there are several worthwhile directions in which to continue. For example, the texts studied in this dissertation represent only one item in the genre chain of documents with which a literary scholar might engage. For example, the May 2019 issue of *Publications of the Modern Language Association* contains not only a single-novel analysis (“Hunger in the Garden: Shortage and Environmental Aesthetics in Nadine Gordimer’s *The Conservationist*”), but also a film analysis “*Margaret* and the Victorians”, two comparative studies (“Suspicion on Trial: Tolstoy’s *The Kreutzer Sonata* and Nabokov’s ‘*Pozdnyshv’s Address*’” & “Held in Checks: Du Bois, Johnson, and the Figurative Work of Financial Forms”) and a study on poetry (“The Atlantic Ocean Breaking on Our Heads: Claudia Rankine, Robert Lowell, and the Whiteness of the Lyric Subject”). Moreover, there are also recommendations from the Modern Language Association for humanities scholars to become proficient in myriad web genres to make scholarship of greater import and influence toward the public (Purdy & Walker, 2010). Further studies could, then, examine whether any lexical bundles appear specific to distinct kinds of LCAs, noting, for example, if any cluster more often to the communicative purpose of Comparing when placing events from separate novels in dialogue with one another. Looking into linguistic features used in common types of web genres found to be fruitful for publicly engaged literary scholars may prove to be an interesting direction as well.

Additionally, in new light recently shed on the importance of three-word bundles (Cortes, 2019) in academic writing, bundles of this length in both languages warrant examination in future studies. Whereas four-word bundles represented an average of 1.4% of the prose in English and 4% of the prose in a Spanish bundle, the percentage of prose constituted by bundles grew to 13% in Spanish and 5.7% in English when including three-word bundles. A larger number of lexical bundles from which to choose may yield insight into some highly-frequent bundles which reliably perform a wider variety of descriptive communicative purposes.

Next, although the title of this dissertation contains the word “macrostructures”, there remains work to be done to find more comprehensive patterns of repetition in LCW. In line with other software tools that generate paragraph-level scores for linguistic features such as ReaderBench (Dascalu et al., 2015), future studies should investigate variation in dimension scores from paragraph to paragraph in LCAs using, for example, Sharoff’s (2020) Multidimensional Analysis Framework programmed in Python. To illustrate, being able to measure the levels of narrative features in a single paragraph may facilitate the identification of a passage as one that primarily enacts the Retelling communicative purpose. A follow-up visualization tool that allows students to navigate to passages high or low in particular dimensional features could also allow further insight into guided analyses of LCW as they learn to write their own articles. Further, such software could assist in addressing one goal specified in the title of this work: to see the unseen humanities macrostructures.

Finally, more work with participants is also needed. Not only would research in this area benefit from classroom-based studies on the teaching of lexical bundles in literary analysis classrooms in English and Spanish, but it would also benefit from interviews with scholars who

publish in multiple languages (cf. Belcher & Yang, 2020). In that way, the demands made of English- and/or Spanish-speaking literary scholars might be better understood.

7.6 Concluding Remarks

This dissertation examined the feasibility of MDA, LB, and QCA methodologies to provide descriptive data on published LCW in English and Spanish. Although the data is generated from only a single type of literary criticism paper (analysis of a single novel), it provides a tangible departure for the teaching of lexicogrammatical features characteristic of articles in top-tier journals in both languages. Specifically, the dissertation contains a macroscopic view of discourse features showing that LCW differs from both general academic prose and fiction writing in both languages in distinct fashions. The complementary microscopic view using the LB and QCA methods offers clarity on register-specific phrases used to achieve specific communicative purposes in both descriptive and inferential communicative purposes. I suggest that instructors design activities asking students first to ask what communicative purposes are achieved by lexical bundles in passages containing these expressions, before moving on to using them in an article, finally, substituting words in the both with those from a thesaurus to facilitate further eschewing of repetitive language. For example, the corpus contains the phrases used to put forth interpretations *can be seen as, can be read as, can be interpreted as,* etc. Students could be taught to use these bundles, as well as look for synonyms for *to see, to read, or to interpret* that would still make sense in context. The same could be suggested for the Spanish bundle set. These materials are crucial for the facilitation of the scholarly careers of literary critics who work in both languages. For L1 and L2 Spanish users in Spanish degree

programs, producing a decent LCA may constitute part of their success at the graduate level in Spanish. For L1 and L2 English users in the field of Spanish literature, publishing an article in a top-tier journal in English may add a/an (arguably undue) perception of rigorous academic qualifications. Further, as humanities departments remained threatened by deeper budget cuts (Nichols, 2019), I argue that a well-trained cohort of graduate students who applies these skills by engaging at conferences and, indeed, making an impact by publishing in journals may increase a growing number of humanities departments' reputations, thereby allowing them more funding to continue the mission of appealing to students' latent abilities to use works of literature to engage constructively with the societal issues of today and tomorrow.

REFERENCES

- Afros, E., & Schryer, C. F. (2009). Promotional (meta) discourse in research articles in language and literary studies. *English for Specific Purposes*, 28(1), 58 – 68.
- Al-Surmi, M. (2012). Authenticity and TV shows: A multidimensional analysis perspective. *TESOL Quarterly*, 46(4), 671-694.
- Altenberg, B. & Eeg-Olofsson, M. (1990). Phraseology in spoken English: presentation of a project. In D'Arcy, J. (Ed.), *Theory and practice in corpus linguistics*, (pp. 1 – 26). Amsterdam: Rodopi.
- Altenberg, B. (1993). Recurrent word combinations in spoken English. In *Proceedings of the Fifth Nordic Association for English Studies Conference*. (Eds.) Aarts, J. & Meijs, W., 17 – 27. Reykjavik: University of Iceland.
- Álvarez López, F. (2013). *La despersonalización en el discurso académico escrito* (Doctoral dissertation). University of Alcalá.
- Anthony, L. (2014). AntConc (Version 3.4. 3)[Computer Software]. Tokyo, Japan: Waseda University.
- Asención-Delaney, Y. (2014). A Multi-Dimensional analysis of advanced written L2 Spanish. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multidimensional analysis, 25 years on: a tribute to Douglas Biber* (pp. 35 – 79). Amsterdam: John Benjamins.
- Asención-Delaney, Y., & Collentine, J. (2011). A multidimensional analysis of a written L2 Spanish corpus. *Applied linguistics*, 32(3), 299 – 322.
- Askehave, I., & Swales, J. M. (2001). Genre identification and communicative purpose: A problem and a possible solution. *Applied linguistics*, 22(2), 195 – 212.

- Atai, M. R., & Falah, S. (2005). A contrastive genre analysis of result and discussion sections of applied linguistic research articles written by native and non-native English speakers with respect to evaluated entities and ascribed values. *Proceedings of the 10th Conference of Pan-Pacific Association of Applied Linguistics* (pp. 41 – 56). Edinburgh University.
- Aull, L. L., & Swales, J. M. (2015). Genre analysis: Considering the initial reviews. *Journal of English for Academic Purposes*, 19, 6 – 9.
- Beke, R., & Bolívar, A. (2009). Certainty and commitment in the construction of academic knowledge in the humanities. In F. Dervin & E. Suomela-Salmi (Eds.), *Cross linguistic and cross cultural perspectives on academic discourse* (pp. 33 – 48). Amsterdam: John Benjamins.
- Belcher, D. (2007). Seeking acceptance in an English-only research world. *Journal of Second Language Writing*, 16(1), 1 – 22.
- Belcher, D., & Yang, H. S. (2020). Global perspectives on linguacultural variation in academic publishing. *Journal of English for Research Publication Purposes*, 1(1), 28 – 50.
- Berber Sardinha, T., Kauffmann, C., & Acunzo, C. (2014). Dimensions of register variation in Brazilian Portuguese. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multidimensional analysis, 25 years on: a tribute to Douglas Biber* (pp. 35 – 80). Amsterdam: John Benjamins Publishing Company.
- Berber Sardinha, T., & Veirano Pinto, M. (2017). American television and off-screen registers: a corpus-based comparison. *Corpora*, 12(1), 85 – 114.
- Berber Sardinha, T. (2019). Using Multi-Dimensional Analysis to Detect Representations of National Cultures. In T. Berber Sardinha and M. Veirano Pinto (Eds.), *Multi-Dimensional Analysis: Research Methods and Current Issues* (pp. 231 – 258).

- Berger, A. (2011). *Blind men and elephants: Perspectives on humor* (Vol. 1). Piscataway, New Jersey: Transaction Publishers. London: Bloomsbury Academic.
- Bernadini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15 – 38). Amsterdam: John Benjamins.
- Besnier, N. (1988). The linguistic relationships of spoken and written Nukulaelae registers. *Language*, 64(4), 707 – 736.
- Bestgen, Y. (2019). Comparing lexical bundles across corpora of different sizes: the Zipfian problem. Advance online publication. *Journal of Quantitative Linguistics*, 1 – 19.
- Bhatia, V. K. (1997). Introduction: Genre analysis and world Englishes. *World Englishes* 16(3), 313 – 319.
- Bhatia, V. K. (2005). Generic patterns in promotional discourse. In V. K. Bhatia, (Ed.), *Persuasion across Genres: A Linguistic Approach* (pp. 213 – 228). Amsterdam: John Benjamins Publishing Company.
- Biber, D., Johansson, S., Leech, G., Conrad S., & Finegan, E. (1999). *Longman grammar of spoken and written English Harlow*. London: Pearson.
- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora: Studies in honor of Stig Johansson* (pp. 181 – 189). Amsterdam: Rodopi.
- Biber, D., & Conrad, S. (2001). Corpus-Based Research in TESOL: Quantitative Corpus-Based Research: Much More Than Bean Counting. *TESOL Quarterly*, 35(2), 331 – 336.
- Biber, D. (1985). Investigating macroscopic textual variation through multifeature/multidimensional analyses. *Linguistics*, 23(2), 337 – 360.

- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D. (1992). The multi-dimensional approach to linguistic analyses of genre variation: An overview of methodology and findings. *Computers and the Humanities*, 26(5-6), 331 – 345.
- Biber, D. (1993). Using register-diversified corpora for general language studies. *Computational linguistics*, 19(2), 219 – 241.
- Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of Corpora: Studies in Honor of Stig Johansson*. (pp. 181 – 190). Amsterdam: Rodopi.
- Biber, D. (2000). FixTag (Version 2.0) [Computer Software]. Flagstaff, Arizona: Northern Arizona University.
- Biber, D. (2006). Spanish Tagger [Computer Software]. Flagstaff, Arizona: Northern Arizona University.
- Biber, D. (2006). Spanish TagCount [Computer Software]. Flagstaff, Arizona: Northern Arizona University.
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2 – 20.
- Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge: Cambridge University Press.

- Biber, D., & Hared, M. (1992). Dimensions of register variation in Somali. *Language Variation and Change*, 4(1), 41 – 75.
- Biber, D., Conrad, S., & Cortes, V. (2003). Lexical bundles in speech and writing: An initial taxonomy. In A. Wilson., P. Rayson, & T. McEnery (Eds.), *Corpus linguistics by the lune: A festschrift for Geoffrey Leech* (pp. 71 – 92). Bern, Switzerland: Peter Lang.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371 – 405.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25, 371 – 405.
- Biber, D., Davies, M., Jones, J. K., & Tracy-Ventura, N. (2006). Spoken and written register variation in Spanish: A multi-dimensional analysis. *Corpora*, 1(1), 1 – 37.
- Biber, D., Egbert, J., & Davies, M. (2015). Exploring the composition of the searchable web: a corpus-based taxonomy of web registers. *Corpora*, 10(1), 11 – 45.
- Biber, D., & Finegan, E. (1994). Intra-textual variation within medical research articles. *Corpus-based research into language*, 12, 201 – 222.
- Biber, D., & Gray, B. (2013). Discourse characteristics of writing and speaking task types on the TOEFL x® test: a lexico-grammatical analysis. *ETS Research Report Series*, 2013(1), i – 128.
- Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press.
- Bitchener, J. (2010). *Writing an applied linguistics thesis or dissertation: A guide to presenting empirical research*. Basingstoke, England: Palgrave Macmillan.

- Blackwell, F. H., & Larson, P. E. (2007). *Guía básica de la crítica literaria y el trabajo de investigación*. Boston, MA, USA: Thomson Higher Education.
- Booth, W. C. (2010). *The rhetoric of fiction*. Chicago: University of Chicago Press.
- Brinkschulte, M., Stoian, M. E., & Borges, E. (2015). Resource-focused Research for Multilingual Competence in Scientific Writing. *Journal of Academic Writing*, 5(1), 17 – 28.
- Bruce, I. (2008). Cognitive genre structures in Methods sections of research articles: A corpus study. *Journal of English for Academic Purposes*, 7(1), 38 – 54.
- Butler, C. S. (1998). Collocational frameworks in Spanish. *International journal of corpus linguistics*, 3(1), 1 – 32.
- Caeiro, C., Cruz, E. B., & Pereira, C. M. (2014). Arts, literature and reflective writing as educational strategies to promote narrative reasoning capabilities among physiotherapy students. *Physiotherapy theory and practice*, 30(8), 572 – 580.
- Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education. *Corpora*, 14(3). Advance online publication.
- Cao, Y., & Xiao, R. (2013). A multi-dimensional contrastive study of English abstracts by native and non-native writers. *Corpora*, 8(2), 209 – 234.
- Chen, L. (2019). *Lexical Bundles in Vocabulary-based Discourse Units: A Corpus-based Study of First Year Engineering Textbooks* (Doctoral dissertation). Carleton University.
- Chen, Y. H., & Baker, P. (2010). Lexical Bundles in L1 and L2 Academic Writing. *Language Learning and Technology*, 14(2), 30 – 49.
- Connor, U., & Upton, T. (2003). Linguistic dimensions of direct mail letters. In *Corpus Analysis* (pp. 71-86). Brill Rodopi.

- Cortes, V. (2002). *Lexical bundles in academic writing in history and biology*. (Doctoral dissertation). Northern Arizona University.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for specific purposes*, 23(4), 397 – 423.
- Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, 17(4), 391 – 406.
- Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora*, 3(1), 43 – 57.
- Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for academic purposes*, 12(1), 33 – 43.
- Cortes, V. (2015). Situating lexical bundles in the formulaic language spectrum: origins and functional analysis development. In V. Cortes & E. Csomay (Eds.), *Corpus-based research in applied linguistics: Studies in honor of Doug Biber*, (pp. 197 – 218). Amsterdam: John Benjamins.
- Cortes, V., & Hardy, D. (2013). Analyzing the semantic prosody and semantic preference of lexical bundles. In D. Belcher & G. Nelson (Eds.), *Critical and corpus-based approaches to intercultural rhetoric*, (pp. 180 – 201). Ann Arbor: University of Michigan Press.
- Cotos, E., Huffman, S., & Link, S. (2017). A move/step model for methods sections: Demonstrating rigour and credibility. *English for Specific Purposes*, 46, 90 – 106.
- Crossley, S. A., Allen, L. K., & McNamara, D. S. (2014a). A multi-dimensional analysis of essay writing. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multidimensional analysis, 25 years on: a tribute to Douglas Biber* (pp. 197 – 238). Amsterdam: John Benjamins.

- Crossley, S. A., Allen, L. K., Kyle, K., & McNamara, D. S. (2014b). Analyzing discourse processing using a simple natural language processing tool. *Discourse Processes*, 51(5-6), 511-534.
- Csomay, E. (2006). Academic talk in American university classrooms: crossing the boundaries of oral-literate discourse? *Journal of English for Academic Purposes*, 5(2), 117 – 135.
- Darhower, M. (2014). Literary discussions and advanced-superior speaking functions in the undergraduate language program. *Hispania*, 97(3), 396 – 412.
- Dascalu, M., Trausan-Matu, S., McNamara, D. S., & Dessus, P. (2015). ReaderBench—Automated Evaluation of Collaboration on the basis of Cohesion and Dialogism. *International Journal of Computer-Supported Collaborative Learning*, 10, 395 – 423.
- Day, R. A. (2005). *Cómo escribir y publicar trabajos científicos* (Vol. 598). Pan American Health Organization.
- Doddington, G. (2002). Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. *Proceedings of the second international conference on Human Language Technology Research* (pp. 138 – 145). San Francisco: Morgan Kaufman Publishers.
- Doyle, M. S. (2018). Spanish for the Professions and Specific Purposes: Curricular Mainstay. *Hispania*, 100(5), 95 – 101.
- Engeli, L., Moergeli, H., Binder, M., Drabe, N., Meier, C., Buechi, S. & Jenewein, J. (2016). Resilience in patients and spouses faced with malignant melanoma. A qualitative longitudinal study. *European journal of cancer care*, 25(1), 122 – 131.
- Fahnestock, J. (1993). Genre and rhetorical craft. *Research in the Teaching of English*, 27(3), 265 – 271.

- Fahnestock, J., & Secor, M. (1991). The rhetoric of literary criticism. In C. Bazerman & J. Paradis (Eds.), *Textual dynamics of the professions: Historical and contemporary studies of writing in professional communities* (pp. 77 – 96). Madison: University of Wisconsin Press.
- Firth, J. R. (1957). *Papers in Linguistics*. Oxford: Oxford University Press.
- Fischer, H., Heise, L., Heinz, M., Moebius, K., & Koehler, T. (2014). E-Learning Trends and Hypes in Academic Teaching. Methodology and Findings of a Trend Study. International Association for Development of the Information Society.
- Francomano, E. C. (2012). Graduate students in Spanish need to become humanities professors. *Hispania*, 95(3), xviii – xix.
- Freedman, A. (1993a). Show and tell? The role of explicit teaching in the learning of new genres. *Research in the Teaching of English*, 27(3), 222 – 251.
- Freedman, A. (1993b). Situating genre: A rejoinder. *Research in the Teaching of English*, 27(3), 272 – 281.
- Friginal, E. (2013). Twenty-five years of Biber's Multi-Dimensional Analysis: introduction to the special issue and an interview with Douglas Biber. *Corpora*, 8(2), 137 – 152.
- Frye, N. (1957). *Anatomy of Criticism: Four Essays*. Princeton: Princeton University Press.
- Fuentes Cortés, M. (2012). El discurso científico de la historia: Análisis estructural y retórico de los artículos de investigación en historia. *Boletín de filología*, 47(1), 89 – 110.
- García, P., & Asención, Y. (2001). Interlanguage development of Spanish learners: Comprehension, production, and interaction. *Canadian Modern Language Review*, 57(3), 377 – 401.

- Gardner, S., Nesi, H., & Biber, D. (2019). Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing. *Applied Linguistics*, 40(4), 646 – 674.
- Geisler, C. (2003). Gender-based variation in nineteenth-century English letter writing. In *Corpus Analysis* (pp. 87 – 106). Brill Rodopi.
- Gilmore, A., & Millar, N. (2018). The language of civil engineering research articles: A corpus-based approach. *English for Specific Purposes*, 51, 1 – 17.
- Gill, C. J. (1996). A Year of Residence in the Household of a South Carolina Planter: Teacher, Daughters, Mistress, and Slaves. *The South Carolina Historical Magazine*, 97(4), 293-309.
- Girgensohn, K. (2016). Students' Writing Research as a Tool for Learning—Insights into a Seminar with Research-Based Learning. *Journal of Academic Writing*, 6(1), 73 – 83.
- Given, L. M., & Willson, R. (2015). Collaboration, information seeking, and technology use: A critical examination of humanities scholars' research practices. In P. Hansen, C. Shah, & C. Klas (Eds.), *Collaborative information seeking: Best practices, new domains and new thoughts* (pp. 139 – 164). New York: Springer.
- Göbel, K., & Helmke, A. (2010). Intercultural learning in English as foreign language instruction: The importance of teachers' intercultural experience and the usefulness of precise instructional directives. *Teaching and Teacher Education*, 26(8), 1571 – 1582.
- González Arías, C. (2011). La formulación de los objetivos en artículos de investigación científica en cuatro disciplinas: Historia, Lingüística, Literatura y Biología. *Linguagem em (Dis)curso*, 11(2), 401 – 429.

- Grafmiller, J. (2014). Variation in English genitives across modality and genres. *English Language & Linguistics*, 18(3), 471 – 496.
- Grant, R. M., & Tracy, D. (1984). *A Short History of the Interpretation of the Bible*. Fortress Press.
- Gray, B. (2015). *Linguistic variation in research articles: When discipline tells only part of the story*. Philadelphia: John Benjamins Publishing Company.
- Gray, C.M. (2010). FixTag (Version 0.5) [Computer Software]. Flagstaff, Arizona: Northern Arizona University.
- Gunkel, P., Weidlich, J., & Bastiaens, T. (2018, March). A Case Study in Flipped Tutoring Development, Implementation and Evaluation of a Flipped Classroom Approach to English Language Tutoring. In E. Langran & J. Borup (Eds.), *Society for Information Technology & Teacher Education International Conference* (pp. 840 – 845). Chesapeake, Virginia: Association for the Advancement of Computing in Education.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Hardy, J. A., & Römer, U. (2013). Revealing disciplinary variation in student writing: A multi-dimensional analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, 8(2), 183 – 207.
- Haswell, R. (1991). *Gaining ground in college writing: Tales of development and interpretation*. Dallas: Southern Methodist University Press.
- Heider, U., & Steger, F. (2014). Individuelle Entscheidungsfindung nach pränatal diagnostizierter schwerer fetaler Fehlbildung. *Ethik in der Medizin*, 26(4), 269 – 285.
- Henning, S.D. (1993). Integration of language, literature and culture: Goals and curricular design. *ADFL Bulletin*, 24(2), 22 – 29.

- Hertel, T. J., & Dings, A. (2014). The undergraduate Spanish major curriculum: Realities and faculty perceptions. *Foreign Language Annals*, 47(3), 546 – 568.
- Hertel, T. J., & Dings, A. (2017). The undergraduate Spanish major curriculum: Faculty, alumni, and student perceptions. *Foreign Language Annals*, 50(4), 697 – 716.
- Hirano, E. (2009). Research article introductions in English for specific purposes: A comparison between Brazilian Portuguese and English. *English for specific purposes*, 28(4), 240 – 250.
- Hoffjan, A., & Wömpener, A. (2006). Comparative analysis of strategic management accounting in German-and English-language general management accounting textbooks. *Schmalenbach Business Review*, 58(3), 234 – 258.
- Hoffman, F. J. (1957). *Freudianism and the literary mind*. Baton Rouge: Louisiana State University Press.
- Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied linguistics*, 20(3), 341 – 367.
- Iglič, H., Doreian, P., Kronegger, L., & Ferligoj, A. (2017). With whom do researchers collaborate and why? *Scientometrics*, 112(1), 153 – 174.
- Jalali, H. (2014). A Comparative Study of Anticipatory it Lexical Bundles in Applied Linguistics and Analytical Chemistry Research Articles. *Journal of Foreign Language Teaching and Translation Studies*, 3(1), 25 – 42.
- Jespersen, O. (1933). *Essentials of English Grammar*. New York: Henry Holt & Co.
- Jian, H. (2010). The Schematic Structure of Literature Review in Research Articles of Applied Linguistics. *Chinese Journal of Applied Linguistics (Foreign Language Teaching & Research Press)*, 33(5), 15 – 27.

- Johansson, S, Leech, G., & Goodluck, H. (1978). *Manual of information to accompany the Lancaster-Oslo/Bergen Corpus of British English, for use with digital computers*. Oslo, Norway: University of Oslo, Department of English.
- Johansson, S. (1982). *Computer corpora in English language research*. Bergen: Norwegian Computing Centre for the Humanities.
- Johansson, S. (2007). *Seeing Through Multilingual Corpora: On the Use of Corpora in Contrastive Studies*. Philadelphia: John Benjamins Publishing.
- Johns, A. M. (1997). *Text, Role and Context: Developing Academic Literacies*. Cambridge: Cambridge University Press.
- Johnston, K. M. (2017). *Lexical Bundles in Applied Linguistics and Literature Writing: A Comparison of Intermediate English Learners and Professionals*. (Master's thesis). Portland State University
- Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for specific purposes*, 24(3), 269 – 292.
- Kanoksilapatham, B. (2007). Rhetorical moves in biochemistry research articles. In D. Biber, U. Connor, & T. A. Upton (Eds.), *Discourse on the move: Using corpus analysis to describe discourse structure* (pp. 73 – 119). Amsterdam: John Benjamins.
- Kanoksilapatham, B. (2015). Distinguishing textual features characterizing structural variation in research articles across three engineering sub-discipline corpora. *English for Specific Purposes*, 37, 74 – 86.
- Karlen, Y., & Compagnoni, M. (2017). Implicit theory of writing ability: Relationship to metacognitive strategy knowledge and strategy use in academic writing. *Psychology Learning & Teaching*, 16(1), 47 – 63.

- Khany, R., & Tazik, K. (2010). A comparative study of introduction and discussion sections of sub-disciplines of applied linguistics research articles. *Research in Applied Linguistics*, 1(2), 97 – 122.
- Kim, Y.-J ., & Biber, D. (1994) A corpus based analysis of register variation in Korean. In D. Biber and E. Finegan (Eds.) *Sociolinguistic Perspectives on Register* (pp. 157 – 81). Oxford: Oxford University Press.
- King, P. (2007). Estudio multidimensional de la oralidad a partir de los textos escolares para la enseñanza del inglés como lengua extranjera. *Revista Signos*, 40(63), 101 – 126.
<https://dx.doi.org/10.4067/S0718-09342007000100006>
- Klee, K.A. (2015). Content-based programs and Spanish for the professions. In M. Lacorte (Ed.), *The Routledge handbook of Hispanic applied linguistics* (pp. 184 – 201). New York: Routledge.
- Kreft, A. & Viebrock, B. (2014). To Read or not to Read: Does a Suitcase Full of Books do the Trick in the English Language Classroom? *CLELE Journal*, 2(1), 72 – 91.
- Krippendorff, K. (1980). *Content analysis. An introduction to its methodology*. Beverly Hills: Sage.
- Kusch, C. (2016). *Literary analysis: The basics*. London: Routledge.
- Kwatsha, L. L. (2007). A psychoanalytical interpretation of the characters in AC Jordan's novel *Ingqumbo yeminyanya*. *Literator: Journal of Literary Criticism, Comparative Linguistics and Literary Studies*, 28(3), 75 – 89.
- Lake, W., & Cortes, V. (2019). Toward a framework for published literary research in Spanish and English for the 2019. Paper presented by William Michael Lake at the American Council for the Teaching of Foreign Languages 2019 Conference, Washington, DC.

- Lake, W., & Cortes, V. (2020). "Using lexical bundles to reassess disciplinary norms in Spanish & English humanities and social science writing". In U. Römer, V. Cortes, E. Friginal (Eds.), *Advances in corpus-based research on academic writing*. (pp. 184 – 203). Amsterdam: John Benjamins.
- Lake, W., & Cortes, V., (in press). "Comparing lexical bundles in Spanish and English psychology methods sections". In E. Friginal & J. Hardy (Eds.), *Routledge handbook of corpus approaches to discourse analysis*. (Accepted). Abingdon, UK: Routledge.
- Larivière, V., Gingras, Y., & Archambault, É. (2006). Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. *Scientometrics*, 68(3), 519 – 533.
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS and R*. London: Routledge.
- Lázaro, C. F., & Correa, C. E. (1980). *Cómo se comenta un texto literario*. Madrid: Cátedra.
- Li, C., & Ngai, J. (2018). Challenges and issues in academic writing: Perspectives from peer tutors. *The Asian Journal of Applied Linguistics*, 5(1), 104 – 116.
- Lide, F. (1988). On Constituting and Institutionalizing a Foreign Language Discipline: The Example of Rhetoric and Composition in Departments of English. *The Modern Language Journal*, 72(1), 42 – 51.
- Liebrenz, M., Fisher, C., Nellen, R., Frei, A., Biechl, A., Hiestand, N., & Eich, D. (2016). Adult attention-deficit/hyperactivity disorder and nicotine withdrawal: a qualitative study of patient perceptions. *BMC psychiatry*, 16(1), 1 – 12.

- Liebrenz, M., Gehring, M. T., Buadze, A., & Caflisch, C. (2015). High-dose benzodiazepine dependence: a qualitative study of patients' perception on cessation and withdrawal. *PLoS one*, *10*(11), 1 – 23.
- Lin, L., & Evans, S. (2012). Structural patterns in empirical research articles: A cross-disciplinary study. *English for Specific Purposes*, *31*(3), 150 – 160.
- Lloyd, P. M. (1987). *From Latin to Spanish: Historical phonology and morphology of the Spanish language* (Vol. 173). American Philosophical Society.
- Loan, N. T. T., & Pramoolsook, I. (2015). Move Analysis of Results-Discussion Chapters in TESOL Master's Theses Written by Vietnamese Students. *3L: Southeast Asian Journal of English Language Studies*, *21*(2), 1 – 15.
- Lowney, J. (2015). "A New Kind of Music": Jazz Improvisation and the Diasporic Dissonance of Paule Marshall's *The Fisher King*. *MELUS: Multi-Ethnic Literature of the United States*, *40*(1), 99 – 123.
- MacDonald, S. (1994). *Professional academic writing in the humanities and social sciences*. Carbondale, IL: SIU Press.
- Maier, V., & Budke, A. (2016). The use of planning in English and German (NRW) Geography school textbooks. *Review of International Geographical Education Online*, *6*(1), 8 – 31.
- Mansilla, M., & Vélez, C. (2016). La pampa asesinada en la ciudad la muerte en el libro *Hijo Del Salitre* de Volodia Teitelboim. *Revista de Letras*, *56*(1), 135 – 150.
- Martín-Martín, P. (2005). *The rhetoric of the abstract in English and Spanish scientific discourse: A cross-cultural genre-analytic approach*. Bern, Switzerland: Peter Lang
- Mayring, P. (2014). *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Klagenfurt, Austria: Gesis: Leibniz-Institut für Sozialwissenschaften.

- Mayring, P. (2015). Qualitative content analysis: Theoretical background and procedures. In A. Bikner-Ahsbahr, C. Knipping & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education* (pp. 365 – 380). New York: Springer.
- Mbodj, N., & Crossley, S. (2020a). “Students’ use of lexical bundles: Exploring the discipline and writing experience interface”. In U. Römer, V. Cortes & E. Friginal (Eds.), *Advances in corpus-based research on academic writing*. (pp. 116 – 133). Amsterdam: John Benjamins.
- Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis*. Thousand Oaks, California: Sage.
- Modern Language Association. (2007). Foreign languages and higher education: New structures for a changed world. Retrieved March 10, 2011, from <http://www.mla.org/flreport>.
- Moreno, A. I., & Swales, J. M. (2018). Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes*, 50, 40 – 63.
- Mur Dueñas, P. (2009). Logical markers in L1 (Spanish and English) and L2 (English) business research articles. *English Text Construction*, 2(2), 246 – 264.
- Nance, K.A. (2015). Hispanic literatures and cultures throughout the curriculum. In M. Lacorte (Ed.), *The Routledge handbook of Hispanic applied linguistics* (pp. 238 – 257). New York: Routledge.
- Narbona, A. (2000). Sintaxis coloquial. In M. Alvar (Ed.), *Introducción a la lingüística española* (pp. 463 – 476). Barcelona: Ariel.
- Navarro, F. (2014). *Manual de escritura para las carreras de Humanidades*. Buenos Aires: Facultad de Filosofía y Letras Universidad de Buenos Aires

- Nichols, W. (2019). Telling the Story of Iberian Studies: Spaces of Convergence and the Defense of the Humanities. Paper presented at the 2019 Modern Language Association Convention, Chicago, IL.
- Nodoushan, M. A. S., & Khakbaz, N. (2011). Theses 'Discussion' sections: A structural move analysis. *International Journal of English Studies*, 5(3), 111 – 132.
- Oermann, M. H., Nicoll, L. H., Chinn, P. L., Ashton, K. S., Conklin, J. L., Edie, A. H., ... & Williams, B. L. (2018). Quality of articles published in predatory nursing journals. *Nursing Outlook*, 66(1), 4 – 10.
- Otañi, I. (2005). El resumen o abstract del artículo de investigación en español: Características retórico-lingüísticas. In C. Bruck (Ed.), *Español con fines académicos: De la comprensión a la producción de textos*, (pp. 45 – 72) Madrid: Editorial Edinumen.
- Palmer, H. E. (1933). Second interim report on English collocations. In the *Tenth Annual Conference of English Teachers under the Auspices of the Institute for Research in English Teaching*. Tokyo, Institute for Research in English Teaching, 1933.
- Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60 – 71.
- Parodi, G. (2005). Lingüística de corpus y análisis multidimensional: Exploración de la variación en el Corpus PUCV-2003. in G. Parodi, (Ed.), *Discurso especializado e instituciones formadoras*. (pp. 83 – 125) Valparaíso: Ediciones Universitarias de Valparaíso.
- Parodi, G. (2007). Variation across registers in Spanish: exploring the El Grial PUCV corpus. In G. Parodi (Ed.), *Working with Spanish Corpora* (pp. 11 – 53). London: Continuum.

- Parodi, G. (2010). Multisemiosis y lingüística de corpus: Artefactos (multi) semióticos en los textos de seis disciplinas en el corpus PUCV-2010. *RLA. Revista de lingüística teórica y aplicada*, 48(2), 33 – 70.
- Payant, C., and J. A. Hardy. 2016. The dynamic rhetorical structures of TESOL conference abstracts. *BC TEAL Journal* 1(1): 1–17.
- Pearson, L., Fonseca-Greber, B., & Foell, K. (2006). Advanced proficiency for foreign language teacher candidates: What can we do to help them achieve this goal? *Foreign Language Annals*, 39(3), 507 – 519.
- Perales Escudero, M., & Swales, J. M. (2011). Tracing convergence and divergence in pairs of Spanish and English research article abstracts: The case of Ibérica. *Ibérica*, 21, 49 – 70.
- Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, 14, 84 – 94.
- Pérez-Paredes, P., & Sánchez-Tornel, M. (2019). The linguistic dimension of L2 interviews: A multidimensional analysis of native speaker language. *Focus on ELT Journal*, 1(1), 4 – 26.
- Petrovich, E. (2018). Accumulation of knowledge in para-scientific areas: the case of analytic philosophy. *Scientometrics*, 116(2), 1123 – 1151.
- Pharies, D. A. (2015). *Breve historia de la lengua española: Segunda edición revisada*. University of Chicago Press.
- Picoral, A. (2018). Biber Tag Checking Tool (Version 1.0) [Computer Software]. Tuscon, Arizona: The University of Arizona.
- Purdy, J. P., & Walker, J. R. (2010). Valuing digital scholarship: Exploring the changing realities of intellectual work. *Profession*, 2010(1), 177 – 195.

- Rafieyan, V. (2018). Knowledge of formulaic sequences as a predictor of language proficiency. *International Journal of Applied Linguistics and English Literature*, 7(2), 64 – 69.
- Reppen, R. (2007). L1 and L2 writing development of elementary students: Two perspectives. In Y. Kawaguchi, T. Takagaki, N. Tomimori, and Y. Tsurugura (Eds.) *Corpus-based perspectives in linguistics Vol. 6*. (pp. 147 – 167). Amsterdam: John Benjamins.
- Robinson, M., Stoller, F., Costanza-Robinson, M., & Jones, J. K. (2008). *Write like a chemist: A guide and resource*. Oxford University Press.
- Rosmarin, A. (1985). *The power of genre*. University of Minnesota Press.
- Ruiying, Y., & Allison, D. (2003). Research articles in applied linguistics: Moving from results to conclusions. *English for specific purposes*, 22(4), 365 – 385.
- Schluer, J. (2014). Writing for publication in linguistics: Exploring niches of multilingual publishing among German linguists. *Journal of English for Academic Purposes*, 16, 1 – 13.
- Shapiro, N.M. (2004). North Alabama's response to the fear and facts of slave revolts. *Huntsville Historical Review*, 30(1), 15 – 54.
- Sharoff, A. (2020). BiberPy. [Computer Software]. Leeds: University of Leeds.
- Sheldon, E. (2011). Rhetorical differences in RA introductions written by English L1 and L2 and Castilian Spanish L1 writers. *Journal of English for Academic Purposes*, 10(4), 238 – 251.
- Sheldon, E. (2018). Dialogic spaces of knowledge construction in research article Conclusion sections written by English L1, English L2 and Spanish L1 writers. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos*, (35), 13 – 40.

- Shin, Y. K., & Kim, Y. (2017). Using lexical bundles to teach articles to L2 English learners of different proficiencies. *System*, 69, 79 – 91.
- Shin, Y. K., Cortes, V., & Yoo, I. W. (2018). Using lexical bundles as a tool to analyze definite article use in L2 academic writing: An exploratory study. *Journal of Second Language Writing*, 39, 29 – 41.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Slinker, B. K., & Glantz, S. A. (1985). Multiple regression for physiological data analysis: the problem of multicollinearity. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 249(1), R1 – R12.
- Small, H. (2015). Hispanic literatures and cultures throughout the curriculum. In R. McDonald (Ed.), *The values of literary studies: Critical institutions, scholarly agendas* (pp. 27 – 43). New York: Routledge.
- Soto, G., & Zenteno, C. (2004). Los sintagmas nominales en textos científicos escritos en español. *Estudios de Lingüística*, 18, 275 – 292.
- Soto, G., Martínez, R., & Sadowsky, S. (2005). Verbos y sustantivos en textos científicos. Análisis de variación en un corpus de textos de ciencias aplicadas, naturales, sociales y humanidades. *Philologia Hispalensis*, 19, 169 – 187.
- Stoller, F. L., Horn, B., Grabe, W., & Robinson, M. S. (2006). Evaluative review in materials development. *Journal of English for Academic Purposes*, 5(3), 174–192.
- Svalberg, A. M. L. (1995). Meanings into pictures: Icons for teaching grammar. *Language Awareness*, 4(2), 65-87.
- Svartvik, J., & Quirk, R. (1980). *A corpus of English conversation*. Lund: CWK Gleerup.

- Swales, J. (1981). *Aspects of article introductions* (Aston ESP research report no. 1). Language Studies Unit, The University of Aston, Birmingham.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Swales, J. M., & Feak, C. B. (2004). *Academic writing for graduate students: Essential tasks and skills* (Vol. 1). Ann Arbor, MI: University of Michigan Press.
- Szeghi, T. M. (2014). Weaving transnational cultural identity through travel and diaspora in Sandra Cisneros's *Caramelo*. *MELUS: Multi-Ethnic Literature of the United States*, 39(4), 162-185.
- Tankó, G. (2017). Literary research article abstracts: An analysis of rhetorical moves and their linguistic realizations. *Journal of English for Academic Purposes*, 27, 42 – 55.
- Terveen, I.C. (2013). Frequency and function of codeswitching among German-English bilingual preschool children in Cape Town. (Doctoral dissertation). Stellenbosch University.
- Thieme, K. (2017). Do We Need New Method Names? Descriptions of Method in Scholarship on Canadian Literature. *ESC: English Studies in Canada*, 44(1), 91 – 110.
- Titak, A., & Roberson, A. (2013). Dimensions of web registers: an exploratory multi-dimensional comparison. *Corpora*, 8(2), 235 – 260.
- Tötösy de Zepetnek, S. (2007). The New Humanities: The Intercultural, the Comparative, and the Interdisciplinary. *The Global South*, 1(1), 45 – 68.
- Tracy-Ventura, N., Cortes, V., & Biber, D. (2007). Lexical bundles in speech and writing. In G. Parodi (Ed.), *Working with Spanish corpora* (pp. 217–230). London, UK: Continuum.
- Tröger, E., Treleaven, N., Lanitz, F., Wendling, C. & Brush, M. (2019). Geany (Version 1.36) [Computer Software].

- Tucker, P. (2014). Evaluation and interpretation in art-historical discourse. In L. Camiciotti, G. Del & E. Tognini (Eds.), *Academic Discourse - New Insights into Evaluation* (pp. 161 – 179). Bern, Switzerland: Peter Lang.
- Turner, J. (2015). *Philology: The forgotten origins of the modern humanities* (Vol. 70). Princeton University Press.
- Ucar, I., López-Fernandino, F., Rodriguez-Ulibarri, P., Sesma-Sanchez, L., Urrea-Micó, V., & Sevilla, J. (2014). Growth in the number of references in engineering journal papers during the 1972–2013 period. *Scientometrics*, 98(3), 1855 – 1864.
- Venegas, R. (2010). Multi-dimensional analysis of an academic corpus in Spanish. In G. Parodi (Ed.), *Academic and professional discourse genres in Spanish* (pp. 101 – 119). Philadelphia, PA: John Benjamins.
- Viera Echevarría, C. (2014). Academic Spanish in the United States: Discourse analysis of academic conference presentations (Doctoral dissertation). University of California, Davis.
- Viera Echevarría, C. (2017). Genre and register variation: Academic conference presentations in Spanish in the United States. In J.J. Colomina-Almiñana (Ed.), *Contemporary advances in theoretical and applied Spanish linguistic variation*, (pp. 148 – 164). Columbus, Ohio: The Ohio State University Press.
- Vine, E. (2011). High frequency multifunctional words: Accuracy of word-class tagging. *Te Reo*, 54, 71 – 82.
- Voigt, A., & Girgensohn, K. (2015). Peer Tutoring in Academic Writing with Non-Native Writers in a German Writing Center—Results of an Empirical Study. *Journal of Academic Writing*, 5(1), 65 – 73.

- Wainer, H. (1992). Understanding graphs and tables. *Educational researcher*, 21(1), 14 – 23.
- Warren, J. E. (2006). Literary scholars processing poetry and constructing arguments. *Written Communication*, 23(2), 202 – 226.
- White, L., Spada, N., Lightbown, P. M., & Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics*, 12(4), 416 – 432.
- Wilder, L. (2005). “The rhetoric of literary criticism” revisited: Mistaken critics, complex contexts, and social justice. *Written Communication*, 22(1), 76 – 119.
- Wilder, L. (2012). *Rhetorical strategies and genre conventions in literary studies: Teaching and writing in the disciplines*. Carbondale, IL: SIU Press.
- Wilder, L., & Wolfe, J. (2009). Sharing the tacit rhetorical knowledge of the literary scholar: The effects of making disciplinary conventions explicit in undergraduate writing about literature courses. *Research in the Teaching of English*, 44(2), 170 – 209.

APPENDICES

Appendix A. English-language Training Document

Seeking the Unseen Humanities Macrostructures Communicative Purposes Taxonomy Training Document

Defining:

A definition of theoretical constructs or other key terms.

Relevant linguistic features:

Authors or infrequent theoretical concepts as the sentence subject,
stative/cognition/communication verbs, additional descriptive language.

- a. While most people only superficially interact with each other, unquestionably aping one another's assumptions and subordinating themselves to **an anonymous mass that Heidegger calls "the they"**—"We take pleasure and enjoy ourselves as they take pleasure; we read, see, and judge about literature and art as they see and judge; likewise we shrink back from the 'great mass' as they shrink back; we find 'shocking' what they find shocking" (164)—being authentic (eigenlicht) involves interpreting the present world in a manner different from that of the general population.

Describing:

A description of a person or a state of affairs, real or imagined.

Relevant linguistic features:

Stative/cognition verbs, descriptions of habitual actions, and verbs in the present or past perfect aspect.

- a. As the narrator **insists**, these characters "had erotic depths, origins, memories, dreams and seizures of melancholy and enthusiasm" (25).
- b. Eliot **is proud** of his job as a middle-man at a mouthwash company, he **espouses** the virtues of the nuclear family, and he **expresses** concern for his neighbours, while Nellie **is** a homemaker who embraces conservative morality and traditional gender roles.

- c. Others, most vocally Evan McKenzie, **believe** that HOAs (homeowner associations) have made residential neighbourhoods into “privatopias” that corrupt the potential for community.

Retelling:

Refers to a simple restating of information, whether from the author’s biographical details, research by other scholars, or events in taking place in the novel itself. It co-occurs predominantly with the historical present, but the use of the preterite (in Spanish) is an option as well. This action is similar to “describing”, but uses activity and achievement verbs rather than stative or cognition verbs. When retelling examples from the novel itself, it may co-occur with or without the support of direct quotes.

Relevant linguistic features:

The historical present.

- a. When Tony **goes** one step further and **belittles** the mouthwash business, Eliot **swings** his club at his son.
- b. Eliot’s actions **send** Tony into a deep depression, leaving him mysteriously bedridden for months.
- c. Tony’s odd condition, combined with Eliot’s realization that he could have killed his son, **forces** him to recognize the possibility of death.

Explaining:

A presentation of information with some degree of inference that would not appear immediately obvious to the reader.

Relevant linguistic features:

Stative, communication and activity verbs in dialogue with their object-collocates; logical connectives; generalizations.

- a. **Central** to these debates **is the language of western liberalism**, specifically that of social contract theory.
- b. In this arrangement, however, **what one “knows” is not the authentic self but rather the performative persona** dictated by the terms of the contract.
- c. **Cheever illustrates Tony’s growing alterity** in a scene in which he and Eliot go golfing.

Evaluating

Evaluations which may come across as more subjective than experts would normally allow in other disciplines.

Relevant linguistic features:

Impersonal generalizations; words expressing deficit(e.g., “lacks”); “more...than”; attributive adjectives expressing affect, i.e., “compelling;”, pronominal adjectives expressing affect (a notable phenomenon in Spanish).

- a. This emphasis on the monstrous and shallow **indicates that many writers cast the blame for suburbia’s failures**, not on the residential model, but on the actual people who live there.
- b. However, Hoover was not interested in **mere shelter and certainly did not intend to increase public housing**.
- c. The surrealism Cheever employs in Bullet Park **undermines the romance of this authenticity, revealing it to be too singular** and self-absorbed for any sustainable form of community.

Comparing

Can be used to establish an interplay of similarities between:

- a. Characters
- b. elements in the plot
- c. other scholars
- d. other works of fiction
- e. the theoretical framework

Discourse markers may help signal the presence of a comparison, but depending on the author, the reader may become tasked with identifying one sentence as similar to or different from another.

Relevant linguistic features:

Discourse markers indicating comparison and contrast.

- a. Scott Donaldson believes that suburban contracts **misapply the individualist, agrarian spirit of Jefferson and Thoreau**, while Milette Shamir and F.M. Coleman consider the relation between subjectivity and nature in the American transcendentalists a forerunner to suburban land development and contracts.

- b. **Like** his earlier and better known fictional neighbourhood, Shady Hill, Cheever's Bullet Park is a bizarre and pluralistic society inhabited by middle-class oddities.
- c. **Where Nellie recreates her identity in a place she hates, Eliot has** his persona disrupted in the place he loves, his home in Bullet Park.

Appendix B. Spanish language Training Document

Seeking the Unseen Humanities Macrostructures Communicative Purposes Taxonomy Training Document

Defining:

A definition of theoretical constructs or other key terms.

Relevant linguistic features:

Authors or infrequent theoretical concepts as the subject, stative / cognition / communication verb, additional descriptive language.

- b. Para la escritura Roberto Bolaño **tiene una definición rotunda**, que toma de un verso de Nicanor Parra: Es, dice, como ordeñar una vaca para tirarle luego la leche a la cara (Bolaño 2003a: 137).

Describing:

A description of a person or a state of affairs, real or imagined.

Relevant linguistic features:

Stative/cognition verbs, descriptions of habitual actions, and verbs in the present or past perfect aspect.

- d. Cueto **mantiene** la culpabilidad del conserje japonés.
e. Se **sospecha** de un robo.
f. Comprender, interpretar, no **es** solo una instancia científica, escribe Gadamer, sino que pertenece a la experiencia humana del mundo.

Retelling:

Refers to a simple restating of information, whether from the author's biographical details, research by other scholars, or the novel itself. It co-occurs predominantly with the historical present, but the use of the preterite is an option as well. This action is similar to "describing", but uses activity and achievement verbs rather than stative or cognition verbs. When retelling examples from the novel itself, it may co-occur with or without the support of direct quotes.

Relevant linguistic features:

The historical present.

- d. Se **declara** inocente, pero **es detenido**.

- e. En el pueblo **empiezan a circular** distintas versiones de lo ocurrido.
- f. El detective es el que investiga, el que mete la cabeza en lo oscuro, **dice**; nada que ver con una actitud naïf, ni tampoco solo lúdica.

Explaining:

A presentation of information with some degree of inference that would not appear immediately obvious to the reader.

Relevant linguistic features:

Stative, communication and activity verbs in dialogue with their object-collocates; logical connectives; generalizations.

- d. Bolaño **usa el desierto como espacio para esas respuestas**: una metáfora muy visual para la búsqueda durísima, y peligrosa, que es la literatura.
- e. **En esta misma brecha abierta Ricardo Piglia busca** su espacio.
- f. Ricardo Piglia, con Blanco nocturno (2010), una novela policiaca vuelta del revés, **cuestiona el artificio de este género** para reflexionar, en un ejercicio metaliterario que queda superpuesto a la trama, sobre las claves de la interpretación.

Evaluating

Evaluations which may come across as more subjective than experts would normally allow in other disciplines.

Relevant linguistic features:

Impersonal generalizations; words expressing deficit(e.g., “lacks”); “more...than”; attributive adjectives expressing affect, i.e., “compelling;”, pronominal adjectives expressing affect.

- d. **Habría que** inventar un nuevo género policial, la ficción paranoica.
- e. **Es lo más sensato**: la hermenéutica, como filosofía, como epistemología, atenta sobre todo al método, tiene que ser concienzudamente cauta con la articulación de sus propuestas.
- f. **Como si con el molde de la novela policiaca, con las simetrías que genera el método, Piglia quisiera llevar** sus pautas a la órbita de la hermenéutica, con su sugerencia sobre la diversidad de significados, de interpretaciones.

Comparing

Can be used to establish an interplay of similarities between:

- f. Characters
- g. elements in the plot
- h. other scholars
- i. other works of fiction
- j. and the theoretical framework

Discourse markers may help signal the presence of a comparison, but depending on the author, the reader may become tasked with identifying one sentence as similar to or different from another.

Relevant linguistic features :

Discourse markers indicating comparison and contrast.

- d. Convencido de que en la novela pueden expresarse también pensamientos, **aunque de otro modo que en filosofía.**
- e. Con Croce, su protagonista en Blanco nocturno, Piglia se planta ante la filosofía, ante los modos prestigiados del pensamiento, ante el método, **que tuvo con Hegel su apoteosis.**
- f. Pero Piglia, al hacer de la investigación de su comisario algo rudimentario, poco profesional, más intuitivo que racional, **hace con la novela policiaca lo que Cervantes hace con la novela de caballería:** Replantearla desde dentro, cuestionando su propio mecanismo, hasta hacerla colapsar.

Appendix C. Reference chart to aid participants' judgments

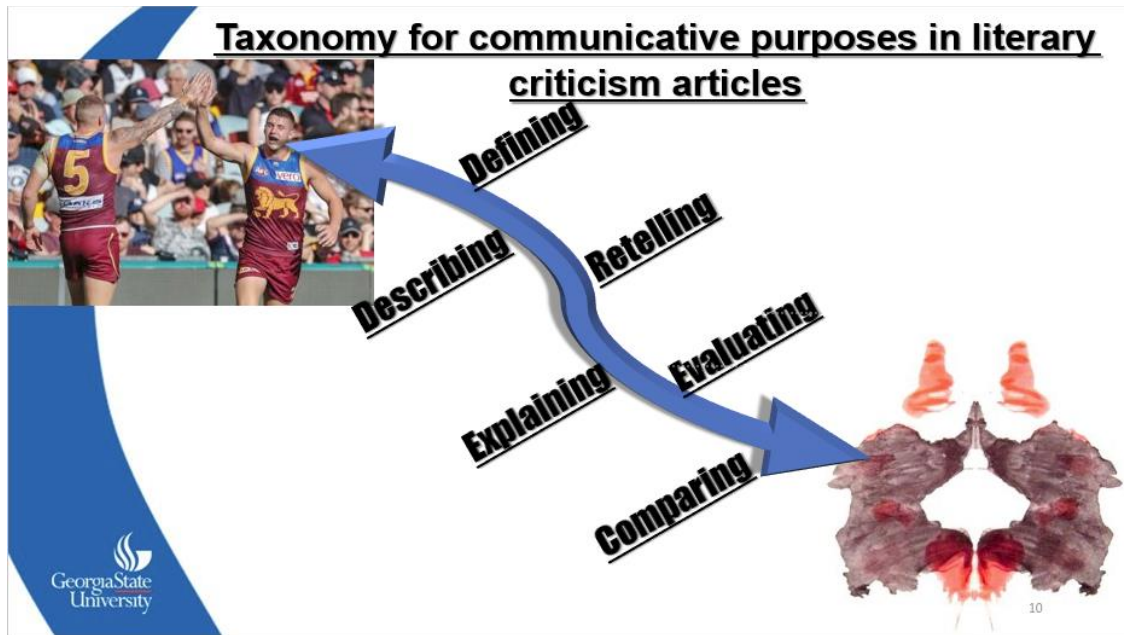


Figure C.1. Reference chart to aid participants' judgments.

Appendix D. Practice Worksheet – English

Seeking the Unseen Humanities Macrostructures: Communicative Purposes Taxonomy Training Document

1. Since its establishment in the waning years of World War II, the US post-war suburb has connoted inauthenticity and conformity.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

2. Anyone familiar with suburban fiction will recognize the teenager’s jeremiad.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

3. The emphasis on conformity further led to exclusion and materialism, as possessions—particularly, the house and yard—and similarity to one’s neighbours were considered the constituent elements of one’s identity.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

4. Due in part to the coalition’s urging, the Federal Government approved a number of mandates designed to encourage the growth of suburbia, just as Levitt & Sons began working on the first modern suburb: Levittown in Long Island, New York.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

5. According to Abigail Cheever’s Real Phonies, this concept of authenticity was particularly prevalent in the post-war period in which the suburbs were born.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

6. Like Heidegger, these thinkers understand authenticity to be, not an isolated substance, but inherently social—based and founded on interruptions, interactions, and relations.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

7. The self is always in relation, and these relations are necessary because no one creates a world according to his or her own wishes, but each of us is “thrown” into an existence already populated with subjects, objects, and moods.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

8. For Heidegger, the death of another person serves as a powerful catalyst for authenticity, as it brings individuals into a condition he calls “Being-toward-death” (Sein-zum-Tode).

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

9. Being-toward-death forces a person to accept the limits of available possibilities, putting greater emphasis on the immediate over the ideal.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

10. Heidegger does not consider this anxiety a bad thing, but rather a significant call to authenticity; when one accepts that he or she can never outstrip the possibility of death, one begins to anticipate it, thereby becoming “resolute.”

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

11. As opposed to the isolated authenticity described by Rousseau and inferred by modern suburban contracts, Heidegger’s version of authenticity is inherently relational and imagines others not as potential contaminants who must be mediated (if not eradicated) but rather necessary elements of identity construction.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

12. Upon entering the Nailles’s house, he asserts his control over it, locking Eliot and Nellie out of their son’s room and insisting he not be disturbed, and he begins his work by cleaning Tony’s room.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

Appendix E. Practice Worksheet – Spanish

Seeking the Unseen Humanities Macrostructures Communicative Purposes Taxonomy Training Document

1. Tras las fases de aculturación y desculturación, surge la neoculturación, en que emergen nuevos fenómenos culturales:

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

2. Al mirarse por dentro, cualquier hermenéutica empieza por justificarse sus límites, qué acepta y qué no acepta para la interpretación, qué puede permitirse el método que propone y qué no.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

3. Como si Piglia, al colapsar la novela policiaca, intentara un nuevo acuerdo de mínimos para la comprensión tras el descreimiento de una sola posible lectura, de la infalibilidad de la lógica, del razonamiento analítico.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

5. Indico rápidamente dos (también contemporáneas, también en español), que han sabido acertar con un símbolo inagotable: Sobre héroes y tumbas (1961), de Sabato, y Los detectives salvajes (1998) del propio Bolaño: ambas sobre las posibilidades epistémicas de la novela: sobre su tema y su recorrido o su acceso: Ernesto Sabato, con una analogía, señala el tema de la literatura: el mal, que representa con los ciegos: un primer punto.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

6. Roberto Bolaño, con una analogía, señala el camino inclemente de la búsqueda: el desierto: un segundo punto.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

7. Piglia aprieta la trama de Blanco nocturno: Tony Durán, un puertorriqueño criado en Nueva Jersey, aparece muerto en su habitación, en el hotel en que se hospeda desde que llegó al pueblo, en la provincia de Buenos Aires.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

8. Se piensa en las gemelas Belladona, en Ada y Sofía, con las que él mantenía un juego a tres bandas.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

9. Son las hijas del viejo Belladona, para el que Tony traía los dólares: un modo de ingresar divisas de modo ilegal: la última posibilidad de salvar la fábrica que ahora lleva Luca.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

10. Croce es cesado e internado en un psiquiátrico.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

11. El criminal ya no es un individuo aislado, sino una gavilla que tiene el poder absoluto.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

12. Aunque Croce sin caer en el relativismo exasperante de Leon, en su desconfianza por hacerse con la realidad: “De todos modos la realidad circundante se hallaba ya contagiada por la posibilidad de distintos significados y esto me separaba, me aislaba de todo”, dice en Cosmos (Gombrowicz 1982: 46), muy lejos del comisario de Piglia, por esa actitud derrotista al final.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

12. En ambos funciona lo que podríamos denominar un mecanismo de dilatación espacial.

Defining / Describing / Retelling / Explaining / Evaluating / Comparing

Appendix F. Four-word LBs from English corpus

Table F.1. English 4-word bundles

Bundle	
as a result of	the fact that the
as part of a	the figure of the
as part of the	the history of the
at the end of	the image of the
at the expense of	the manner in which
at the hands of	the relationship between the
at the heart of	the rest of the
at the same time	the role of the
by the fact that	the way in which
can be seen as	the ways in which
despite the fact that	the world of the
draws attention to the	through the lens of
in a way that	to the extent that
in an attempt to	to the fact that
in an interview with	turns out to be
in relation to the	ways in which the
in terms of the	
in the case of	
in the context of	
in the face of	
in the form of	
in the midst of	
in the name of	
in the novel is	
in the process of	
in the united states	
is a form of	
is one of the	
is part of the	
it is important to	
of the novel is	
one of the most	
that there is a	
the end of the	
the extent to which	
the face of the	

Appendix G. Five-word LBs from English corpus

Table G.2. English 6-word bundles

Bundle
as a result of the
at the beginning of the
at the end of the
at the heart of the
at the same time as
at the same time that
by the end of the
can be read as a
from the perspective of a
in the aftermath of the
in the case of the
in the context of the
in the face of the
in the form of a
in the wake of the
it is worth noting that
of the ways in which
on the level of the
over the course of the
the end of the novel
the second half of the
the way in which the
toward(s) the end of the

Appendix H. 6+-word LBs from English corpus

Table H.3. English 6+-word bundles

Bundle
Chapel Hill: Univ. of North Carolina Press
at the end of the novel
at the turn of the century
by the end of the novel
in the first half of the
in the second half of the
it is important to note that
references are to this edition and will be cited parenthetically in the text as
this is not to say that

Appendix I. Four-word LBs from Spanish corpus

Table I.4. Spanish 4-word bundles

Bundle		
a la hora de	de la novela de	en la casa de(l)
a la luz de(l)	de la novela en	en la descripción de
a la manera de	de la novela es	en la forma de(l)
a la que se	de la novela histórica	en la historia de
a la vez que	de la novela que	en la narrativa de/del
a lo largo de	de la novela se	en la novela de
a partir de esta / estas / este / estos	de la novela y	en la novela el/las/los/la
a partir de la(s)/los	de la obra de	en la novela que
a partir de su/s	de la publicación de	en la novela se
a partir de un/a	de la vida de	en la obra de
a pesar de la(s)	de las novelas de(l)	en la primera parte
a pesar de que	de lo que se	en la que el/los
a pesar de su	de los personajes de	en la que la/las
a propósito de la/las/los	de los personajes que	en la que se
a sí mismo(s) como	de una serie de	en la segunda parte
a través de la/las/los	desde la perspectiva de(l)	en la vida de
a través de su(s)	el caso de la/las/los	en las que el/la/las
a través de un/a	el final de la/los	en las que se
al final de la	el hecho de que	en los que se
al mismo tiempo que	el modo en que	en medio de la
cada uno de los	el momento en que	en medio de un/a
con el fin de	el mundo de la(s)/los	en relación con el/la
con el objetivo de	el problema de la	en torno a la
con la idea de(l)	el punto de vista	en una de las / en uno de los
con respecto a la/las/los	el tema de la/s	en una serie de
de la casa de(l)	el uso de la / las / los	es el caso de
de la década de(l)	en el caso de	es una de las / es uno de los
de la escritura de	en el contexto de	hasta el punto de
de la historia de	en el marco de(l)	la búsqueda de la
de la historia en	en el momento de	la búsqueda de un/a
de la historia que	en el proceso de	la ciudad de México
de la historia y	en el que el	la construcción de un/a
de la muerte de(l)	en el que la	la cuestión de la
de la novela como	en el que se	la descripción de la/las
	en el seno de	la forma de un/a
	en el sentido de(l)	la forma en que
	en el siglo xix	la historia de la
	en el texto de	la historia y la

la idea de la/las/los

Table I.1. (Continued)

la idea de un/a

la imagen de la

la imagen de un/a

la importancia de la/las/los

la mayoría de los/las/la

la posibilidad de una/un

la presencia de la

la presencia de un/a

la problemática de la

la representación de la / las /
los

la vida de la / las / los

la vida y la / las / los / el

los personajes de la

parte de la novela

por medio de la

por medio de un/a

por parte de la/los

que en la novela

se convierte en el/la

se convierte en un/a

su relación con la(s)/el/los

una y otra vez

Appendix J. Five-word LBs from Spanish corpus

Table J.5. Spanish 5-word bundles

Bundle	en el marco de la(s)
a la luz de la(s)	en el marco de un / a
a lo largo de este / esta / esto	en el momento en que
a lo largo de la / los	en el mundo de la(s) / los
a lo largo de su	en el plano de la
a lo largo del relato	en el sentido de que
a lo largo del siglo	en la construcción de la
a lo largo del texto	en la construcción de un/a
a partir de la cual / los cuales / las cuales	en la década de los
a partir de lo que	en la historia de la
a partir del cual se	en la medida en que
a pesar de que el/la	en la primera parte de
al final de la novela	es el caso de la
como en el caso de	la forma en la que
como si se tratara de	la historia de la literatura
de la ciudad de México	la idea de que la / el
de la década de los	la lectura de la novela
de la década de los cincuenta	la primera mitad del siglo
de la historia de la	la segunda mitad del siglo
de la misma manera que	lo que es lo mismo
de la muerte de su	no cabe duda de que
de la novela de la	no es más que un
desde el punto de vista	que se convierte en un/a
desde este punto de vista	se da cuenta de que
desde la perspectiva de la(s) /los	
desde un punto de vista	
el beso de la mujer	
el final de la novela	
el hecho de que la / el / las / los	
el título de la novela	
El / la protagonista de la novela	
en el ámbito de la(s) / los	
en el campo de la / los	
en el caso de la / las / los	
en el contexto de la(s) / los	

Appendix K. 6+-word LBs from Spanish corpus

Table K.6. Spanish 6+-word bundles

Bundle
a finales del siglo xix y
a lo largo de la historia
a lo largo de la novela
a lo largo de toda la novela
a lo largo de toda la / todo el
a lo largo de toda su
a partir de la década de
de los personajes de la novela
del siglo xix y principios del
desde el punto de vista de
desde el punto de vista de la/las/los
el diccionario de la real academia española
en el caso de la novela
en el hecho de que el/la/los
en El siglo de las luces
en la medida en que el/la
en la medida en que no
en la medida en que se
en la primera parte de la novela
en la segunda parte de la novela
la primera parte de la novela
las primeras décadas del siglo xx
se convierte en una especie de
si bien es cierto que el

Appendix L. All English LBs by functional category

Table L.7. All English LBs by functional category

Classification	Bundles
Referential Deixis	references are to this edition and will be cited parenthetically in the text as
Referential Intangible Framing Attributes	as a result of, as a result of the, as part of a, as part of the, at the expense of, at the hands of, at the heart of, at the heart of the, from the perspective of a, in an attempt to, in an interview with, in relation to the, in terms of the, in the case of, in the case of the, in the context of, in the context of the, in the face of, in the face of the, in the name of, in the process of, of the ways in which, on the level of the, the extent to which, the figure of the, the history of the, the image of the, the manner in which, the relationship between the, the role of the, the way in which, the way in which the, the ways in which, to the extent that, ways in which the
Multi-function	in a way that, in the midst of, the end of the
Obligation / directive	it is important to, it is important to note that
Place	in the united states
Quantifiers	the rest of the
Epistemic Impersonal	by the fact that, can be read as a, can be seen as a, despite the fact that, draws attention to the, in the novel is, is a form of, it is worth noting that, of the novel is, that there is a, the fact that the, this is not to say that, through the lens of, to the fact that, turns out to be
Tangible Framing Attributes	(Chapel Hill: Univ. of North Carolina press,, in the form of, in the form of a, the world of the
Time-Event	at the beginning of the, at the end of, at the end of the, at the end of the novel, at the same time, at the same time as, at the same time that, at the turn of the century, by the end of the, by the end of the novel, in the aftermath of the, in the first half of the, in the second half of the, in the wake of the, over the course of the, the end of the novel, the face of the, the second half of the, toward(s) the end of the
Topic Elaboration / Clarification	as a form of, as a kind of, as a means of, as a symbol of, as one of the, as well as a, as well as his, as well as the, as well as to, in contrast to the, in such a way as to, in the same way that, is beyond the scope of this, on the one hand, on the other hand, to come to terms with the, what it means to be
Topic Introduction / Focus	call(s) attention to the, is one of the, is part of the, one of the most

Appendix M. All Spanish LBs by functional category

Table M.8. All Spanish LBs by functional category

Classification	Bundles
Epistemic impersonal	como si se tratara de, de la novela es, el hecho de que, el hecho de que la/el/las/los, en el hecho de que el/la/los, lo que es lo mismo, no cabe duda de que, no es más que un, que se convierte en un/a, se convierte en el/la, se convierte en un/a, se convierte en una especie de, se da cuenta de que, si bien es cierto que el
Intangible Framing Attributes	a la luz de la(s), a la luz de(l), a la manera de, a partir del cual se , a pesar de la(s), a pesar de que, a pesar de que el/la, a pesar de su, a propósito de la/las/los, como en el caso de, con el fin de, con el objetivo de, con la idea de(l), con respecto a la/las/los, de la ciudad de méxico, de la escritura de, de la historia de, de la historia de la, de la historia en, de la historia que, de la historia y, de la muerte de su, de la muerte de(l), de la obra de, de la publicación de, de la vida de, desde el punto de vista, desde el punto de vista de, desde el punto de vista de la/las/los, desde este punto de vista, desde la perspectiva de la(s)/los, desde la perspectiva de(l), desde un punto de vista, el mundo de la(s)/los, el problema de la, el punto de vista, el tema de la/s, el título de la novela, el uso de la/las/los, en el ámbito de la(s)/los, en el campo de la/los, en el caso de, en el caso de la novela, en el caso de la/las/los, en el contexto de, en el contexto de la(s)/los, en el marco de la(s), en el marco de un/a, en el marco de(l), en el momento en que, en el mundo de la(s)/los, en el plano de la, en el proceso de, en el seno de, en el sentido de que, en el sentido de(l), en la casa de(l), en la construcción de la, en la construcción de un/a, en la descripción de, en la forma de(l), en la historia de, en la historia de la, en la medida en que, en la medida en que el/la, en la medida en que no , en la medida en que se, en la narrativa de/del, en la vida de, en relación con el/la, es el caso de, es el caso de la, la búsqueda de la, la búsqueda de un/a, la construcción de un/a, la cuestión de la, la descripción de la/las, la forma de un/a, la forma en que, la historia de la, la historia de la literatura, la historia y la, la idea de la/las/los, la idea de que la/el, la idea de un/a, la imagen de la, la imagen de un/a, la importancia de la/las/los, la lectura de la novela, la posibilidad de una/un, la presencia de la, la presencia de un/a, la problemática de la, la representación de la/las/los, la vida de la/las/los, la vida y la/las/los/el, los personajes de la, parte de la novela, por parte de la/los, que en la novela, su relación con la(s)/el/los, el caso de la/las/los
Multi- Function	a lo largo de, a lo largo de este/esta/esto, a lo largo de la/los, a lo largo de su, a lo largo de toda la/todo el, a lo largo de toda su, a partir de esta/estas/este/estos, a partir de la cual/los cuales/las cuales, a partir de la década de, a partir de la(s)/los, a partir de lo que, a partir de su/s, a partir de un/a, a través de la/las/los, a través de su(s), a través de un/a, el modo en que, en medio de la, en medio de un/a, por medio de la, por medio de un/a
Place- Institution	en el siglo xix, la ciudad de méxico

Quantifiers	la mayoría de los/las/la
Structural Only	a la que se, de lo que se, en el que el, en el que la, en el que se, en la que el/los, en la que la/las, en la que se, en las que el/a/las, en las que se, en los que se

Table M.1. (Continued)

Tangible Framing Attributes	a sí mismo(s) como, de la casa de(l), de la novela como, de la novela de, de la novela de la, de la novela en, de la novela histórica, de la novela que, de la novela se, de la novela y, de las novelas de(l), de los personajes de, de los personajes de la novela, de los personajes que, el beso de la mujer, el diccionario de la real academia española, el/la protagonista de la novela, en El siglo de las luces, en el texto de, en la novela de, en la novela el/las/los/la, en la novela que, en la novela se, en la obra de, la forma en la que
Time-Event	a finales del siglo xix y, a la hora de, a la vez que, a lo largo de la historia, a lo largo de la novela, a lo largo de toda la novela, a lo largo del relato, a lo largo del siglo, a lo largo del texto, al final de la, al final de la novela, al mismo tiempo que, de la década de los, de la década de los cincuenta, de la década de(l), del siglo xix y principios del, el final de la novela, el final de la/los, el momento en que, en el momento de, en la década de los, en la primera parte, en la primera parte de, en la primera parte de la novela, en la segunda parte, en la segunda parte de la novela, hasta el punto de, la primera mitad del siglo, la primera parte de la novela, la segunda mitad del siglo, las primeras décadas del siglo xx, una y otra vez
Topic Elaboration/Clarification	al fin y al cabo, al igual que el/la/las/los, como una de las/como uno de los, en cuanto a la/s, la manera en que, la manera en que el/la/las/los, más allá de la/las/los, no deja de ser, no se trata de, no se trata de un/a, por lo que se refiere a, que a su vez, que se refiere a, que tiene que ver con el/la/las, tiene que ver con el/la, y por lo tanto
Topic Introduction /Focus	cada uno de los, de la misma manera que, de una serie de, de uno de los, en lo que concierne a la, en lo que se refiere a, en torno a la, en una de las/en uno de los, en una serie de, es una de las/es uno de los, por el hecho de, que se trata de un/a, que se trata de un/a/o, se refiere a la(s), se trata de la/las/los, se trata de un/a

Appendix N. Potentially English-specific LBs by functional category

Table N.9. Potentially English-specific LBs by functional category

Function	Bundle
	as a result of
	as a result of the
	as part of a
	as part of the
	at the expense of
	at the hands of
	at the heart of
	at the heart of the
	in an attempt to
	in an interview with
	in terms of the
	in the case of the
	in the face of
Intangible Framing Attributes	in the face of the
	in the name of
	of the ways in which
	in a way that
Multi-Function	in the midst of
Obligation / Directive	it is important to
	it is important to note that
Place Reference	in the united states
	by the fact that
	despite the fact that
Stance Epistemic Impersonal	in the novel is
	it is worth noting that
	through the lens of
Tangible Framing Attributes	in the form of
	in the form of a
	at the end of
	at the same time
	in the aftermath of the
Time-Event	in the wake of the
	as a form of

Function	Bundle
	as a kind of
	as a means of
	as a symbol of
	as well as a
	as well as his
	as well as the
	as well as to
	in contrast to the
	in such a way as to
	in the same way that
Topic Elaboration / Clarification	on the one hand
	on the other hand
	to come to terms with the

Appendix O. Potentially Spanish-specific LBs by functional category

Table O.10. Potentially Spanish-specific bundles by functional category

Function	Bundle
Identification / Focus	en torno a la
	en una serie de
	de la misma manera que
	de una serie de
	en una de las / en uno de los
	a la luz de la(s)
	a la luz de(l)
	a la manera de
	a partir del cual se
	a propósito de la/las/los
	con el fin de
	con el objetivo de
	con la idea de(l)
	con respecto a la/las/los
	de la ciudad de México
	de la escritura de
	de la historia de
	de la historia de la
	de la historia en
	de la historia que
	de la historia y
	de la muerte de su
	de la muerte de(l)
	de la obra de
	de la publicación de
	de la vida de
	desde el punto de vista
desde este punto de vista	
desde un punto de vista	
en el ámbito de la(s)/los	
en el campo de la/los	
en el caso de la novela	
en el caso de la/las/los	
Intangible	en el marco de la(s)
Framing	en el marco de un/a
Attributes	en el marco de(l)
	en el momento en que
	en el mundo de la(s)/los
	en el seno de
	en el sentido de que
	en el sentido de(l)
	en la casa de(l)
	en la construcción de la
	en la construcción de un/a
	en la descripción de
	en la forma de(l)
	en la historia de
	en la historia de la
	en la medida en que no
	en la narrativa de/del
	en la vida de
	por parte de la/los
	a lo largo de
	a lo largo de este/esta/esto
	a lo largo de la/los
	a lo largo de su
	a lo largo de toda la/todo el
	a lo largo de toda su
	a partir de esta / estas / este / estos
	a partir de la cual / los cuales / las cuales
	a partir de la década de
	a partir de la(s) / los
	a partir de lo que
	a partir de su/s
	a partir de un/a
	a través de la / las / los
	a través de su(s)
	a través de un/a
	en medio de la
	en medio de un/a
Multi-Function	por medio de la
	por medio de un/a

Function	Bundle
Place- Institution	en el siglo xix
Referential Tangible	en El siglo de las luces
	a la que se
	de lo que se
	en el que el
	en el que la
	en el que se
	en la que el/los
	en la que la/las
	en la que se
	en las que el/a/las
Structural Only	en las que se
	en los que se
	a sí mismo(s) como
	de la casa de(l)
	de la novela como
	de la novela de
	de la novela de la
	de la novela en
	de la novela histórica
	de la novela que
	de la novela se
	de la novela y
	de las novelas de(l)
	de los personajes de
	de los personajes de la novela
	de los personajes que
	en el texto de
	en la novela de
	en la novela el / las / los / la
	en la novela que
Tangible Framing Attributes	en la novela se
	en la obra de
	a la hora de
Time-Event	a lo largo de la novela

Function	Bundle
	a lo largo de toda la novela
	a lo largo del relato
	a lo largo del siglo
	a lo largo del texto
	de la década de los
	de la década de los cincuenta
	de la década de(l)
	en el momento de
	en la década de los
	en la primera parte
	en la primera parte de la novela
	hasta el punto de
Topic Elaboration- Clarification	y por lo tanto
	por lo que se refiere a
	de uno de los
	en lo que concierne a la
Topic Introduction / Focus	en lo que se refiere a
	por el hecho de

Appendix P. Heat map of Participants' categorizations: English

Brighter shades of blue indicate a denser concentration of sentences categorized according to a given category.

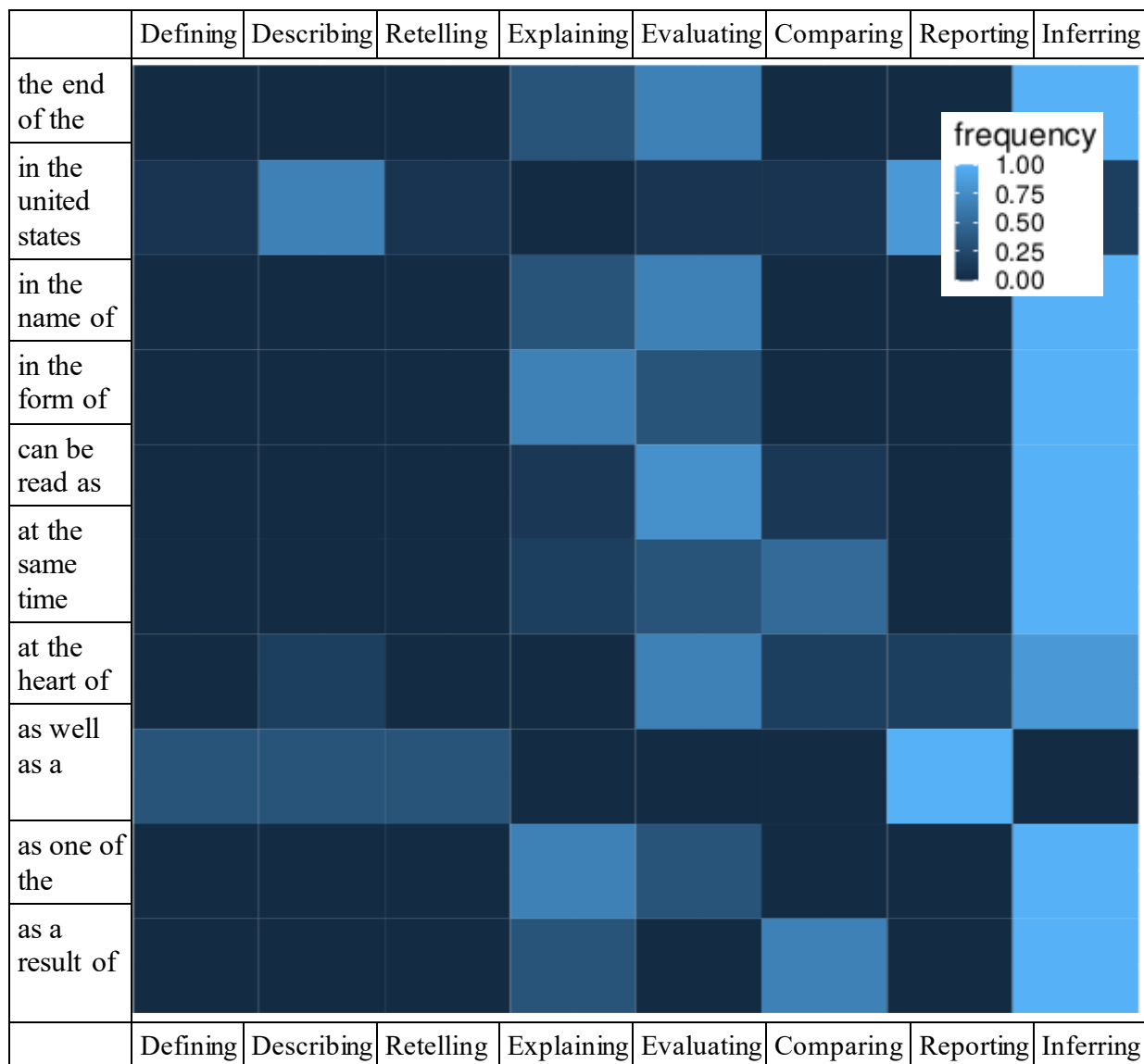


Figure P.2. Heat map of participants' categorizations of LB data in Spanish.

Appendix Q. Heat map of Raters' categorizations: Spanish

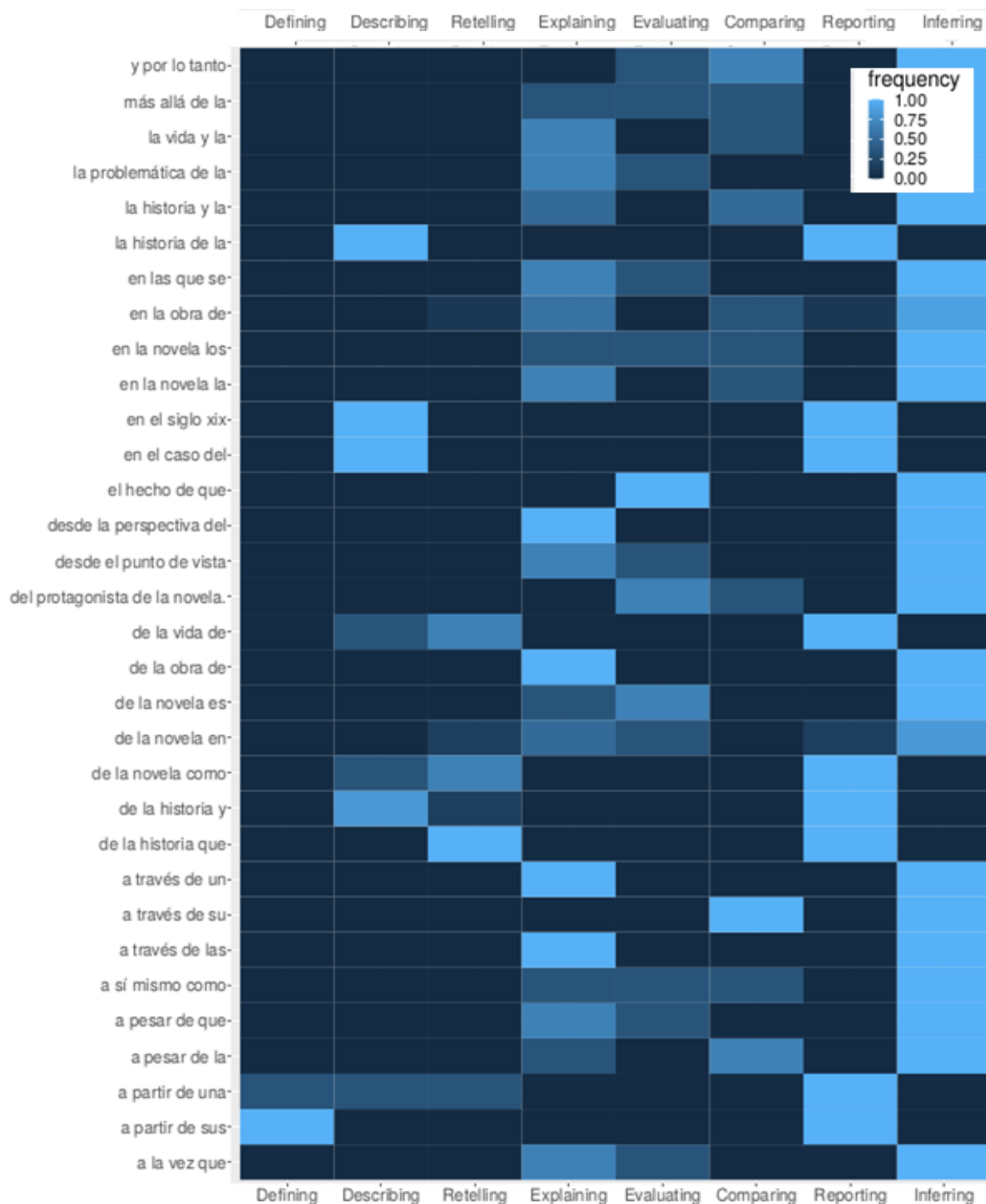


Figure Q.3. Heat map of participants' categorizations of LB data in Spanish.

Appendix R. Informed Consent Form

Georgia State University

Informed Consent

Title: Seeking the unseen humanities macrostructures: The use of corpus- and genre-assisted research methodologies to analyze written norms in English and Spanish literary criticism research articles

Principal Investigator: Viviana Cortes

Student Principal Investigator: William Michael Lake

Sponsor: The National Honor Society of Phi Kappa Phi.

Procedures

You are being asked to take part in a research study whose goal is to understand more deeply how published literary criticism articles are written in English and Spanish. If you decide to take part, you will attend an hour-long training online in which you be trained to rate sentences, in English or Spanish depending on your field of graduate study, for their relative degrees of subjectivity vs. objectivity. After your training, you will read sentences from 5-6 passages from 5-6 different literary criticism articles. You will rate each sentence contained in a spreadsheet on a scale from 1 (objective) to 6 (subjective), along with a short note on why you think so in the adjacent spreadsheet cell. The English-language data set consists of 52 sentences. Completing it should take you approximately 90 minutes. Participants working with Spanish language data will be assigned 106 sentences, which should take you about 3 hours and 20 minutes.

Compensation

The English-language data set consists of 52 sentences. Completing it should take you approximately 90 minutes, for which you will earn \$83.33. Participants working with Spanish language data will be assigned 106 sentences, which should take you about 3 hours and 20 minutes, for which you will earn \$166.67 as a consequence of the higher workload.

Voluntary Participation and Withdrawal

You do not have to be in this study. You may skip questions or stop participating at any time. However, withdrawal will also mean forgoing the above-mentioned compensation.

Contact Information

If you have concerns about this study at any time, please contact Viviana Cortes at (404) 413-5195 (E-mail: vcortes@gsu.edu) or William Michael Lake at (404) 413-5983 (E-mail: wlake1@gsu.edu).

Consent

If you are willing to volunteer for this research, please sign below:
