

A STAIRCASE MODEL FOR TEACHING
GRAMMAR FOR EAP WRITING IN THE IEP:
FRESHMAN COMPOSITION AND THE NOUN
PHRASE

by

GENA R. BENNETT

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Department of English
School of English, Drama, and
American and Canadian Studies
The University of Birmingham
May 2011

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

ABSTRACT

The interface of corpus linguistics and second language writing has led to extensive corpus-based research focusing on a description of academic writing. The overwhelming majority of this research, however, has focused on scholarly writing, which may not be a valid model for novice writing. This thesis proposes the teaching of second language writing should be informed by a staircase model of writing progression which aims instruction at the level of student writing. For English for academic purposes writers in intensive English programs, this target is first year undergraduate writing, specifically freshman composition as it is taught in North American higher education contexts. This study specifically compares the frequency of the noun phrase in freshman composition writing and scholarly writing with two main aims: to provide empirical evidence of the differences between the two levels of writing and to contribute to a description of freshman composition writing. The findings from this comparison clearly demonstrate that noun phrases in both levels of writing employ a discernible pattern, and there are distinct differences between those patterns. A critical need for pedagogical materials to focus more on phrasal structures in general, but especially noun phrases, is evident.

ACKNOWLEDGEMENTS

A very special thank you is due to many people without whom the completion of this thesis would not have been possible. Thank you first and foremost to God; without His grace and supernatural stamina this thesis would never have been finished on time.

Thank you also to the many kind friends who generously agreed to watch Brayden during the month of April. Thank you to Susan Hunston and Oliver Mason for their guidance on the content and structure of the thesis, particularly for the encouraging comments which helped me to carry on. Thank you to my sister Tina for her careful copying editing and to my sister Lena for her encouraging words and helpful hints for avoiding a nervous breakdown. Thank you to Trina because she is my sister, too. Thank you to my parents who always encouraged a love of learning in and expected the best of me. Thank you to the instructors who collected and sent in students' assignments which were a vital part of the study reported here. Thank you to Brayden for believing that I would finish my work someday and having patience until that time so we could finally play together. And an extra special thank you to Brandy for everything you do, but particularly for believing in me, encouraging me, helping me with the technical aspects of formatting and Excel, and doing literally everything in our lives over the past month so that I could complete this thesis.

TABLE OF CONTENTS

CHAPTER ONE INTRODUCTION

1.1	Brief Overview	1
1.2	Relevant Background	2
	<i>1.2.1 Contexts</i>	<i>2</i>
	1.2.1.1 English for Academic Purposes (EAP)	2
	1.2.1.2 Intensive English Program (IEP)	3
	1.2.1.3 Freshman Composition (freshman comp)	3
	<i>1.2.2 Disciplines</i>	<i>4</i>
	<i>1.2.3 Literature</i>	<i>5</i>
1.3	Current Research	7
	<i>1.3.1 Purposes & Aims</i>	<i>7</i>
	<i>1.3.2 Research Questions</i>	<i>8</i>
	<i>1.3.3 Methods & Materials</i>	<i>8</i>
	<i>1.3.4 Findings & Significance</i>	<i>10</i>
1.4	Thesis Structure	11

CHAPTER TWO RESEARCH CONTEXT

2.1	Introduction	15
2.2	North American Higher Education	16
	<i>2.2.1 Options for post-secondary study</i>	<i>16</i>
	<i>2.2.2 Community and four year colleges</i>	<i>16</i>
	<i>2.2.3 General education courses</i>	<i>19</i>
	<i>2.2.4 Freshman Composition</i>	<i>22</i>
	2.2.4.1 Background	22
	2.2.4.2 Organization	22
	2.2.4.3 Controversies	24
	2.2.4.4 Summary	30
2.3	English for Academic Purposes (EAP) and the Intensive English Program (IEP)	30
	<i>2.3.1 EAP</i>	<i>30</i>
	<i>2.3.2 EAP writing</i>	<i>33</i>
	<i>2.3.3 The IEP</i>	<i>36</i>
2.4	A Staircase Model to Inform EAP Writing Instruction in the IEP	37
	<i>2.4.1 Proposal of the model</i>	<i>37</i>
	<i>2.4.2 Description of the model</i>	<i>41</i>
	<i>2.4.3 Theories informing the model</i>	<i>41</i>
	<i>2.4.4 Freshman comp as a discipline</i>	<i>43</i>
	<i>2.4.5 Ascending the staircase</i>	<i>44</i>
	<i>2.4.6 Potential criticisms of the model</i>	<i>47</i>
	<i>2.4.7 Applicability of the model</i>	<i>49</i>
2.5	Summary	49

**CHAPTER THREE
RESEARCH IN SECOND LANGUAGE WRITING
AND CORPUS LINGUISTICS**

3.1	Introduction	51
3.2	Second Language (L2) Writing Research	52
	3.2.1 <i>History of L2 writing</i>	52
	3.2.2 <i>Theory and methodology in L2 writing</i>	53
	3.2.3 <i>Research trends in L2 writing</i>	54
3.3	Research in Corpus Linguistics	58
	3.3.1 <i>History of L2 writing</i>	58
	3.3.2 <i>Theory and methodology in corpus linguistics</i>	59
	3.3.3 <i>Research trends in corpus linguistics</i>	63
3.4	Corpus-based L2 Writing Research	73
	3.4.1 <i>The need for a corpus-based approach</i>	73
	3.4.2 <i>Genre-based approaches</i>	75
3.5	Research Represented in the Levels of the Staircase Model	77
	3.5.1 <i>Organization</i>	77
	3.5.2 <i>Scholarly writing (step 4)</i>	77
	3.5.3 <i>Graduate writing (step 3)</i>	83
	3.5.4 <i>Upper-level undergraduate writing (step 2b)</i>	86
	3.5.5 <i>First year undergraduate writing (step 2a)</i>	89
	3.5.6 <i>Learner writing</i>	91
	3.5.7 <i>The gap in the research</i>	94
3.6	Summary	96

**CHAPTER FOUR
THE CORPORA: FRESHMAN COMPOSITION WRITING
AND SCHOLARLY WRITING**

4.1	Introduction	98
4.2	Corpus-Cited References	99
	4.2.1 <i>Introduction</i>	99
	4.2.2 <i>The Longman Grammar of Spoken and Written English</i>	100
	4.2.3 <i>Cambridge Grammar of English</i>	102
	4.2.4 <i>A Communicative Grammar of English</i>	103
	4.2.5 <i>Summary</i>	105
4.3	The NAFWiC (North American Freshman Writing Corpus)	105
	4.3.1 <i>Justification</i>	105
	4.3.2 <i>Development</i>	106
	4.3.3 <i>Design and organization</i>	108
	4.3.4 <i>Text types and assignment descriptions</i>	111
	4.3.4.1 Expository Essay Subcorpus	112
	4.3.4.2 Persuasive Essay Subcorpus	113
	4.3.4.3 Personal Narrative Subcorpus	114
	4.3.4.4 Proposing a Solution Essay Subcorpus	114
	4.3.4.5 Summary Synthesis Subcorpus	115
	4.3.4.6 Rhetorical Analysis Subcorpus	116
	4.3.4.7 Remembering an Event Subcorpus	116
	4.3.4.8 Profile Subcorpus	117

4.3.4.9	Film Analysis Subcorpus	117
4.3.4.10	Literary Research Paper Subcorpus	118
4.3.4.11	Thematic Analysis Subcorpus	119
4.3.4.12	Character Analysis Subcorpus	119
4.3.4.13	Assignment Topics	120
4.3.5	<i>Representation</i>	120
4.3.5.1	Freshman Composition	121
4.3.5.1.1	Introduction	121
4.3.5.1.2	Instructor A	122
4.3.5.1.3	Instructor B	124
4.3.5.1.4	Instructor C	125
4.3.5.1.5	Instructor D	127
4.3.5.1.6	Overview	128
4.3.5.2	Size	129
4.3.6	<i>File names and annotation</i>	131
4.3.7	<i>Research and continued development</i>	134
4.4	Summary	135

CHAPTER FIVE IDENTIFYING NOUN PHRASES IN FRESHMAN AND SCHOLARLY WRITING

5.1	Introduction	137
5.2	Methods used for scholarly writing	138
5.2.1	<i>Understanding corpus-cited references</i>	138
5.2.2	<i>Approaching corpus-cited references</i>	143
5.2.2.1	A Systematic Approach	143
5.2.2.2	Salient Features of Scholarly Writing	148
5.2.3	<i>The noun phrase in scholarly writing</i>	155
5.2.4	<i>Summary</i>	158
5.3	Methods used for the NAFWiC	158
5.3.1	<i>Overview</i>	158
5.3.2	<i>Identifying noun phrases</i>	159
5.3.2.1	Modification Patterns	159
5.3.2.2	Definite Article	162
5.3.3	<i>Analyzing noun phrases</i>	163
5.3.3.1	Introduction	163
5.3.3.2	N.P. Raw #	163
5.3.3.3	No Modifier: Raw #	163
5.3.3.4	No Modifier: %	164
5.3.3.5	Pre-Modifier: Raw #	164
5.3.3.6	Pre-Modifier: %	164
5.3.3.7	Post-Modifier: Raw #	164
5.3.3.8	Post-Modifier: %	164
5.3.3.9	Both: Raw #	164
5.3.3.10	Both: %	165
5.3.3.11	<i>the</i> : Raw #	165
5.3.3.12	<i>the</i> : %	165
5.3.3.13	Proper Noun: Raw #	165

5.3.3.14	Proper Noun: Factor	166
5.3.3.15	Pre-M Noun: Raw #	166
5.3.3.16	Pre-M Noun: %	166
5.3.3.17	Pre-M Adj: Raw #	166
5.3.3.18	Pre-M Adj: Raw %	166
5.3.3.19	Pre-M Adj <i>and</i> Adj: Raw #	167
5.3.3.20	Pre-M Adj <i>and</i> Adj: %	167
5.3.3.21	Post-M <i>of</i> : Raw #	167
5.3.3.22	Post-M <i>of</i> : %	167
5.3.3.23	Post-M <i>ot.</i> Prep: Raw #	168
5.3.3.24	Post-M <i>ot.</i> Prep: %	168
5.3.3.25	Post-M <i>app</i> : Raw #	168
5.3.3.26	Post-M <i>ot.</i> Prep: %	168
5.3.3.27	Post-M <i>other</i> : Raw #	169
5.3.3.28	Post-M <i>other</i> : %	169
5.3.3.29	Post-M Rel Cl: Raw #	169
5.3.3.30	Post-M Rel Cl: %	169
5.3.3.31	Omit relativizer: Raw #	169
5.3.3.32	Omit relativizer: %	170
5.3.3.33	<i>Which</i> relativizer: Raw #	170
5.3.3.34	<i>Which</i> relativizer: %	170
5.3.3.35	<i>That</i> relativizer: Raw #	170
5.3.3.36	<i>That</i> relativizer: %	171
5.3.3.37	Data Quality Control	171
5.4	Summary	173

CHAPTER SIX COMPARATIVE RESULTS AND DISCUSSION

6.1	Introduction	174
6.2	Lexical Word Classes	175
6.2.1	<i>Distribution of lexical word classes</i>	175
6.2.2	<i>Discussion</i>	177
6.3	The Noun Phrase	178
6.3.1	<i>The definite article</i>	178
6.3.1.1	Distribution in Scholarly Writing	178
6.3.1.2	Distribution in The NAFWiC	180
6.3.1.3	Discussion	182
6.3.2	<i>Pre-modifiers</i>	183
6.3.2.1	Distribution in Scholarly Writing	183
6.3.2.2	Distribution in The NAFWiC	185
6.3.2.3	Discussion	186
6.3.3	<i>Post-modifiers</i>	187
6.3.3.1	Distribution in Scholarly Writing	187
6.3.3.2	Distribution in The NAFWiC	190
6.3.3.3	Discussion	193
6.3.4	<i>Non and Dual Modification</i>	195
6.3.4.1	Distribution in Scholarly Writing	195
6.3.4.2	Distribution in The NAFWiC	196

6.3.4.3	Discussion	197
6.3.5	<i>Summary</i>	198
6.4	A Schematic Representation of the Noun Phrase	200
6.4.1	<i>Introduction</i>	200
6.4.2	<i>Schematic Representation for the NAFWIC and Scholarly Writing</i>	203
6.4.3	<i>Discussion</i>	204
6.5	Summary	206

CHAPTER SEVEN PEDAGOGICAL APPLICATIONS

7.1	Introduction	208
7.2	Pedagogical Materials	209
7.2.1	<i>Definition</i>	209
7.2.2	<i>Development</i>	209
7.3	Corpus Findings in Pedagogical Materials	212
7.3.1	<i>Prospects</i>	212
7.3.2	<i>Problems</i>	213
7.3.3	<i>Potentials</i>	214
7.4	Applications for Pedagogical Materials	215
7.4.1	<i>Survey</i>	215
7.4.1.1	Writers at Work	217
7.4.1.2	Longman Academic Writing Series	221
7.4.1.3	Great Writing	224
7.4.1.4	Summary	229
7.4.2	<i>Recommendations</i>	232
7.4.2.1	Models	232
7.4.2.2	Text types	232
7.4.2.3	Grammar	234
7.4.2.4	The noun phrase	237
7.5	Summary	243

CHAPTER EIGHT CONCLUSION

8.1	Brief Overview	245
8.2	Restatement of Research Aims	246
8.3	Review of Findings	247
8.4	Significance of Findings	249
8.4.1	<i>For corpus linguistics</i>	249
8.4.2	<i>For L2 writing</i>	250
8.5	Limitations of Study	252
8.6	Further Research	253

REFERENCES	257
-------------------	------------

APPENDICES	300	
<i>Appendix 1</i>	<i>Sample General Education Requirements</i>	300
<i>Appendix 2</i>	<i>NAFWiC Informed Consent Form</i>	307
<i>Appendix 3</i>	<i>QTag Accuracy Analysis</i>	309

<i>Appendix 4</i>	<i>Linguistic Features of Scholarly Writing by Part of Speech</i>	<i>312</i>
<i>Appendix 5</i>	<i>Linguistic Features of Scholarly Writing by Function</i>	<i>322</i>
<i>Appendix 6</i>	<i>Negative Salient Linguistic Features of Scholarly Writing</i>	<i>327</i>
<i>Appendix 7</i>	<i>Excel Worksheets for LITR Subcorpus</i>	<i>329</i>
<i>Appendix 8</i>	<i>CD Rom: Raw Data from the NAFWiC (13 Excel workbooks)</i>	<i>Encl</i>

CHAPTER ONE

INTRODUCTION

1.1 Brief Overview

In teaching writing, instructional models allow potentially excluded participants an opportunity to become involved in academic discourse by facilitating participants' realization of the structures and functions exhibited in the targeted discourse and allowing for eventual independent construction of such discourse through experience and participation. I argue here that expert writing is not an appropriate model for novice writers: purposes of expert and student writing differ; the expert standard is unfair and descriptively inadequate; and the model requires far transfer¹ and thus is ineffective. Yet, a student model may also be problematic: multiple scholars criticize student writing as inauthentic and models of student writing may disregard gatekeepers (Hüttner, 2008). In order to facilitate effective writing instruction, this thesis proposes the teaching of second language writing (L2 writing) should be informed by a staircase model of writing progression which aims instruction at the level of student writing. For English for academic purposes (EAP) writers in intensive English programs (IEPs), this target is first year undergraduate writing, specifically freshman composition as it is taught in North American higher education contexts. This proposal effectively addresses all of the aforementioned concerns, as materials stemming from the proposed model are more likely to present writing that can be achievable for students, reflects their communicative purposes, but is also acceptable to the gatekeepers of the discourse, thereby allowing learners to engage with and grow into disciplinary possibilities and find their own academic voice as authors.

¹ This concept of transfer in learning is discussed in Chapter 2.

To serve as an informant to such materials, this study employed corpus-based methods to investigate the noun phrase in freshman composition and scholarly writing and reported differences between the two levels of writing in a schematic representation; the representation highlights the characteristic features of the noun phrase in first year undergraduate writing. It is these features which should be emphasized for EAP writers in the IEP who are seeking to begin undergraduate studies in North American contexts.

The design of this project is motivated by three concerns: the teaching of L2 writing in EAP, specifically in the IEP; the ubiquitous nature of freshman composition in North American higher education; and research on L2 writing in the English as a second language field, specifically corpus-based studies of academic writing.

1.2 Relevant Background

1.2.1 Contexts

The research—and the applications of it—in this study are intimately connected to three contexts.

1.2.1.1 English for Academic Purposes (EAP)

The study of EAP is concerned with communication skills in English which are required for study purposes in formal education systems. Students may need EAP for higher education studies in their own country (e.g. for reading academic texts) or for higher education in English speaking countries. In the North American context, it is typically students who desire to earn a higher education degree from a college or university in that context—and whose first language is not English—that enroll in non-credit bearing EAP courses in order to obtain English language skills which will allow them to successfully participate in academia. EAP programs are typically understood in terms of local contexts and the needs of particular students; EAP practitioners “find out what the students have to do

and help them to do it better” (Gillett, 1996, p. 1). Courses in an EAP program typically include those on academic writing, oral communication, and intensive reading. For those academic writing courses, instructors focus on processes that are necessary for completing writing tasks, institutional and contextual constraints embedded in those tasks, and linguistic choices which have to be made in order to successfully complete those tasks. In doing so, instructors should concentrate on the development of four aspects: schemata, rhetorical patterns, social awareness, and language.

1.2.1.2 Intensive English Program (IEP)

The Intensive English Program (IEP) is one particular English language teaching context in which EAP is often taught. IEPs are post-secondary programs of instruction in public or private tertiary institutions in the US that are designed to develop and strengthen the English language skills of students whose native language is not English but are preparing to enter North American universities. Typically, IEPs conform to a set of guidelines put forth by accrediting organizations, and a typical IEP curriculum is designed for students at most levels of English language proficiency (absolute beginning levels are often not served in IEP programs), with an emphasis on developing the necessary oral and written skills for academic studies. The principles and teaching practices of EAP writing are consistent with those in the IEP. Thus, the North American teaching context of the IEP is ideal for preparing students for the North American freshman composition course.

1.2.1.3 Freshman Composition (freshman comp)

Freshman composition is a ubiquitous, though controversial and contentious, course in the North American higher education context designed to introduce students to the expectations of college writing and to help improve rhetorical skills to meet those expectations. Although there is great diversity in the freshman composition course from one

institution to the next, or even within a single institution, generally freshman comp involves the study and practice of the process approach to writing in the academic community with a focus on critical thinking and research methods and typically engages students in assignments such as rhetorical analysis, research papers, and literacy narratives. Specialists within the freshman comp field often find themselves questioning the purpose of the course, and therefore, how the course should be taught. Regardless of the controversies and negative connotations surrounding freshman comp, it is a salient feature of the university experience for all undergraduates. As the staircase model for the progression of academic writing skills in higher education proposed in this thesis illustrates, freshman comp is the next step for the English language learner who wishes to embark on an undergraduate education in North America; as such, using the freshman comp course as a focus for academic writing instruction in the IEP can efficiently prepare L2 writers for a successful ascension of the academic writing staircase.

1.2.2 Disciplines

The value of corpus linguistics as a discipline has expanded rapidly in the field of English language teaching in the past decade. The largest collection of corpus research has presented findings on how language may better be understood and described. An important element of these findings includes the actual use of language, such as phraseology and lexicogrammar in different registers of language, though more recent corpus linguistics publications have also focused on applications to language teaching. Similarly, the researching and teaching of second language (L2) writing has, in the past decade, steadily become more prevalent, with researchers from English language teaching, applied linguistics, communication, composition studies, and education identifying themselves as L2 writing specialists. L2 writing research is driven by four categories—writers, writers' texts, readers,

and contexts of writing—with an overwhelming focus on pedagogical implications.

Numerous scholarly books and articles have been published in the last decade specifically on these topics of corpus linguistics and second language writing, many with the intended audience of teacher trainees in graduate courses or as a resource for professionals seeking to continue their professional development in the English language teaching field.

The field of L2 writing was initially heavily influenced by that of first language (L1) writing. Yet, the more scholars began to understand the needs of L2 writers, the less L1 writing theories were able to contribute to an understanding of L2 writing. For example, although not all L1 writers, indeed probably few, maintain an absolute control of linguistic features and functions of language, the vast majority possess sufficient control for grammatical ability not to be a major focus of analysis in or instruction for L1 composition. Perhaps not surprisingly, one of the most noticeable differences between L1 and L2 writing is the role of linguistic competence; “based on a vast body of research, limited vocabulary and grammar are the most frequently cited/noted properties of L2 text” (Hinkel, 2011, p. 529). Additionally, L1 writers may have some understood knowledge of cultural expectations for writing, whereas L2 writers may not only have differing knowledge of expectations for writing, but are often still developing proficiency in the L2, making grammatical form as demanding as content. Thus, L2 writing researchers needed to turn to disciplines more rooted in the study of language. As writing is a multidimensional form of communication that involves control of rhetorical, linguistic, and social conventions (Matsuda 1998), the interface with corpus linguistics is a logical choice.

1.2.3 Identifying gaps in the literature

The interface of corpus linguistics and L2 writing has led to extensive corpus-based research focusing on the academic written register. It is common to find L2 writing research

focusing on measures of fluency, accuracy and complexity or applying corpus-based methods in a range of areas including vocabulary, genre, grammar, or citation practices. As is typical for L2 writing, many of these studies are driven "...by the desire to investigate features of academic writing...[and] to bring those discoveries to the classroom..." (Charles, Pecorari, and Hunston, 2009, p. 6). The overwhelming majority of corpus-based writing research has focused on scholarly writing, that writing which has been published and is written by authors with a graduate or postgraduate education to readers of a technical audience, and to a lesser extent, graduate level writing. These studies are often discipline specific, investigating a wide range of lexicogrammatical features and functions in a wide variety of disciplines; furthermore, these studies typically compare findings across register, discipline, level, and even language. Within the past five years, however, the development of register-specific freely available corpora, such as the MICUSP² or BAWE Corpus, have led to an increasing number of studies focusing on upper-level undergraduate writing. As with scholarly and graduate writing, these studies often involve investigation of features or functions in a particular discipline and/or compare findings across disciplines³. Unlike scholarly and graduate writing, however, there is more variation in the genres studied; this typifies the greater range of genres or text types represented in the university setting. As the MICUSP and BAWE become more widely available, we can expect continued growth in the amount of research focusing on this level of writing.

Although research on upper-level undergraduate writing is increasing, research into first year undergraduate writing is virtually non-existent. Of the (merely) 10 studies on first year undergraduate writing conducted in the last decade, only 20% use a corpus-based approach; that is, only two of the studies have aimed to understand lexicogrammatical

² These corpora are, of course, explained more fully later in this thesis (Chapter 3, specifically).

³ As will be discussed in Chapter 3, however, many of these studies are genre-based.

features which characterize first year undergraduate writing! Because this thesis proposes freshman composition as a model for teaching EAP writing in the IEP, the lack of research focusing on first year undergraduate writing is both striking and lamentable. There is an obvious gap in the research.

1.3 Current Research

1.3.1 Purposes & Aims

This gap in the research makes it difficult to answer the continuing question of what linguistic features and functions should be taught in L2 writing. Numerous studies on learner writing have served to provide a description of L2 writers' grammatical competencies and the features of their texts, but instruction informed solely on the findings from these studies focuses on a deficit model of grammar teaching, addressing weaknesses and fixing errors. This type of instruction, focusing on remediation or general language proficiency, is unlikely by itself to help students succeed in university level academic writing or to be especially motivating as students may perceive it as "more of the same." Though much corpus-based L2 writing research in the past two decades has contributed to a description of the academic written register, as discussed above, the majority of this research has investigated expert writing, which is a poor model for novice writers such as those in the IEP. To address the deficiency of the "expert writing" model and to contribute to a fuller description of first year undergraduate writing, this thesis proposes a staircase model of the progression of academic writing development in higher education to inform English for academic purposes (EAP) writing instruction in the intensive English program (IEP). The staircase model is composed of 4 steps in academic writing which could be labeled expert to novice, though here they have been identified by the context and setting in which they take place—scholarly, graduate, undergraduate, EAP—and advocates teaching writing to those on step 1 of the staircase

(EAP) using a writing model informed by that composed on step 2 of the staircase (freshman composition).

In order to facilitate a natural progression from step 1 to step 2, first year undergraduate writing must be researched so that instructors can be aware of the grammar used in successful freshman comp so that they may ensure students understand it. Because nouns and the noun phrase are integral structures in the written academic register, this study specifically compares the frequency of the noun phrase in freshman composition writing and scholarly writing with two main aims: to provide empirical evidence of the differences between the two levels of writing and to contribute to a description of first year undergraduate writing. Fulfilling both of these aims would, in turn, contribute to more efficient EAP writing instruction in the IEP.

1.3.2 Research Questions

This thesis looks to answer the following questions:

1. Do novice writers in freshman composition use noun phrases differently than scholarly writers in published works? If so, to what extent?
2. How can the findings from the first question be used to improve EAP writing materials in the IEP setting?

1.3.3 Methods & Materials

Corpus analysis, in addition to being approached from a use perspective, is distinguished by four major characteristics (Biber, Conrad, & Reppen, 1998, p. 4): they are empirical, analyzing the actual patterns of language use in natural texts; utilize a large and principled collection of natural texts as the basis for analysis; make extensive use of computers for analysis; and depend on both quantitative and qualitative analytical techniques. Furthermore, two approaches characterize studies in corpus linguistics. Word-based studies

focus on the investigation of words, identifying meanings, patterns, and associations of that word and typically begin at the sentence level, making use of concordance lines to access language in the corpus. Context-based studies typically begin with the investigation of a register (or a context within a register), identifying the combination and application of words and categories which distinguish that context from another, often making use of statistical calculations and corpus annotations. The research reported in this thesis is a context-based corpus analysis which makes use of frequency lists, tagging, and parsing to investigate the frequency of the noun phrase in scholarly writing and first year undergraduate writing.

Two general academic writing corpora informed the description of scholarly writing used in this study. Together, the corpora contain approximately 34 million words of scholarly writing from a variety of academic disciplines. Findings from these corpora have been published extensively in corpus-cited references, extensive grammar and vocabulary books that cite corpus findings. The corpus-cited references utilized here include *The Longman Grammar of Spoken and Written English* (Biber et al., 1999), the *Cambridge Grammar of English* (Carter & McCarthy, 2006), and *A Communicative Grammar of English* (Leech & Svartvik, 2002). These references were used to compile a profile of scholarly writing, identifying linguistic features and functions which characterize it. The North American Freshman Writing Corpus (NAFWiC) was created for and used in this study for the investigation of first year undergraduate writing. The NAFWiC includes nearly 250 authentic writing samples that received a grade of an A or B (merit or distinction equivalent) in freshman composition courses from five institutions across North America. The NAFWiC is organized into 12 subcorpora, each representing one text type, and contains nearly 330,000 tokens.

1.3.4 Findings & Significance

The findings from the investigation of the frequency of the noun phrase in scholarly and first year undergraduate writing were presented in a schematic representation of the two levels of writing. The representations clearly illustrate that noun phrases in both levels of writing employ a discernible pattern and there are distinct differences between the two. Though there are multiple options in the noun phrase—use of one, the other, or both modifiers, and a range of modifier types—the schematic representations succinctly depict to what extent each choice is made and to what extent the writers at each level make different choices. Although percentages will not give information on statistical significance, they provide a useful rough means of differentiating between texts. These percentages plainly confirm a difference in the NAFWiC and scholarly writing. In sum, the findings in this study highlight several patterns which characterize the noun phrase in successful first year composition writing. The NAFWiC

- shows a strong preference for modifying noun phrases;
- makes productive use of prepositional phrases and relative clauses as post modifiers;
- makes prolific use of adjectives as pre-modifiers;
- frequently omits relativizers; and
- prefers the use of *that* as a relativizer.

The findings also indicate general observations which distinguish the NAFWiC from scholarly writing; in comparison to scholarly writing, the NAFWiC maintains a(n)

- overall greater use of modifiers,
- less reliance on the definite article,
- greater use of single adjective pre-modifiers,

- more diverse use of pre- and post-modifier types.

It is worth noting that these results are important as they reveal distinct differences between expert and novice writing that have heretofore gone unnoticed. These substantial differences emphasize the argument in this thesis that scholarly writing is a poor model for EAP writers.

In EAP writing instruction, because learner writers are still developing proficiency in their L2, form is just as demanding as content, and the question is not whether or not to teach grammar but how best to do so. Frequency should play a key role in determining what grammar is taught, but frequency should also be considered in tandem with relevance. Basing the choice of what grammar to teach on the schematic representation of the noun phrase (presented in conjunction with the findings) allows materials writers and teachers to consider both. For example, although appositives are used rather infrequently in academic writing, their inclusion in instructional materials is merited to provide students with explicit exposure to the structure; yet, instruction should also note the infrequency of the structure.

In considering pedagogical applications of these findings, this study also undertakes a survey of the six most popular textbooks used in teaching IEP writing to assess how they address the relevant issues discussed throughout the thesis, namely, the use of student models, a focus on both form and function, and attention to relevant text types. The review revealed a critical need for pedagogical materials to focus more on phrasal structures in general, but especially noun phrases.

1.4 Thesis Structure

This thesis begins by establishing the context of the research discussed here. Chapter 2 describes the institutional frameworks of the North American higher education context, the concept of general education courses there, and the important aspects of the freshman composition course as part of that context. Chapter 2 also explores teaching English for

academic purposes (EAP) in the intensive English program (IEP). Additionally, the relevance of freshman composition to that of L2 writing in IEPs is established, and the staircase model of progression in academic writing skills is proposed as a basis to inform academic writing instruction and materials development in the IEP.

Chapter 3 follows by presenting an overview of research in the fields of L2 writing and corpus linguistics, including a discussion of the debates which surround theory and method in both fields; L2 writing lacks a single, comprehensive theory, and researchers in corpus linguistics can not agree on its status as theory, discipline, or method. Chapter 3 also considers the expansion of research in both fields and the natural convergence of the two as L2 writing research shifted to a focus on describing student texts propelled by the desire to investigate features of academic writing and bring those discoveries to the classroom; this, in turn, has led to studies focusing on descriptions of the written academic register. Specifically, research conducted in the contexts appearing on the staircase model of progression in academic writing skills introduced in Chapter 2—scholarly writing, graduate writing, upper-level undergraduate writing, first year undergraduate writing, and learner writing—is reviewed in order to establish the gap in the research. As mentioned above, with only two studies contributing to a lexicogrammatical description of first year writing, there is an obvious gap in the literature.

Chapter 4 describes the corpora used in the comparison of the noun phrase. Firstly, an overview of two large general academic writing corpora whose findings have been published in corpus-cited references is provided. Findings from these corpora published in the corpus-cited references represent the findings for scholarly writing reported here. Following this, an introduction and description of the NAFWiC, an exemplar corpus representing the discipline-

specific writing of freshman composition is provided. Findings from the NAFWiC represent the findings for first year undergraduate writing reported here.

Chapter 5 details the methods employed to provide empirical evidence of the differences between scholarly writing and freshman composition and to contribute to a description of first year undergraduate writing. Firstly, in providing an overview of corpus-cited references, Chapter 5 describes how the references can be useful to both instructors and researchers. One such example is the scholarly writing profile presented in the chapter, which can be exploited by instructors to inform teaching materials and explored by researchers to identify paths of inquiry. Secondly, Chapter 5 explains how the data used in constructing the modified noun phrase in scholarly writing was extracted from the corpus-cited references, including the methods used to approach the references; following this, the chapter explains how these approaches and the creation of the scholarly writing profile led to the identification of the noun phrase as a point of investigation between scholarly writing and freshman composition. Lastly, Chapter 5 details the methods used in the NAFWiC to identify, classify, and analyze the noun phrase in order to make that comparison.

Chapter 6 reports and discusses the comparative findings of the investigation. Findings regarding distribution of lexical word classes are reported and discussed first, followed by findings for the noun phrase; these findings are reported and discussed according to modification: use of pre-modifier only, post-modifier only, and no or dual modifier⁴. The chapter concludes by presenting a schematic representation of the noun phrase for both levels of writing and discussing the characteristics of noun phrases in the NAFWiC as well as an overall comparison of the noun phrase between the NAFWiC and scholarly writing.

⁴ As noted in Chapter 6, the term “dual modifier” here is used as a succinct way of indicating the presence of both a pre- and post modifier.

Chapter 7 explores the pedagogical applications of the ideas and findings presented throughout the thesis. An operational definition of pedagogical materials is given and issues surrounding the creation and use of materials in the writing classroom are discussed. Issues surrounding the integration of findings from corpus-based studies into such pedagogical materials is also explored. The chapter reports on a survey of current textbooks used in EAP writing classes in the IEP and outlines applications and recommendations for those materials, as well as pedagogical materials in general, based on the proposals and findings presented throughout the thesis.

Chapter 8 is the conclusion of the thesis; this chapter provides a re-statement of the research aims, reviews the findings and their contribution to the fields of corpus linguistics and L2 writing, and discusses further research to continue these contributions.

CHAPTER TWO

RESEARCH CONTEXT

2.1 Introduction

As discussed in chapter 1, the design of this project is motivated by three concerns: how writing in a second language is taught, how corpus descriptions of grammar can be useful to teaching writing, and the ubiquitous nature of freshman composition. An understanding of each of these concerns is essential in order to fully understand the context and implications of this investigation. For this reason, this chapter and the next will focus on the North American higher education context, specifically the course of freshman composition; the state of research on second language (L2) writing in the English as a second language field, specifically that relating to corpus-based studies of academic writing; and the teaching of L2 writing in English for Academic Purposes (EAP), specifically teaching in the Intensive English Program (IEP). The state of research on and the teaching of L2 writing and corpus-based studies will be discussed in Chapter 3. This chapter outlines the essential issues relating to freshman composition and the North American higher education context, as well as the EAP and IEP contexts. Section 2.2 provides an overview of the institutional frameworks of the North American higher education context, discusses the concept of general education courses there, and details important aspects of the freshman composition course¹ as part of that context. Section 2.3 describes the EAP and IEP contexts, and section 2.4 discusses the relevance of freshman composition to the teaching of L2 writing in the IEP, specifically proposing a model for instruction.

¹ The North American *course* is equivalent to the UK *module*.

2.2 North American Higher Education

2.2.1 Options for post-secondary study

After graduation from a North American high school, students have three options in higher education: vocational school, community college, or four-year college/university². Vocational schools often offer courses in professional trades such as hairdressing, auto mechanics, electrical repair, etc. and lead to a certificate. Vocational schools are not pertinent to the current arguments, and, therefore, will not be discussed in any detail. As community colleges and four-year colleges are directly relevant to freshman composition, and as argued here EAP writing, a more complete overview of each, including a description of curriculum, is provided below.

2.2.2 Community and four year colleges

In the United Kingdom, “community” college usually refers to a Sixth Form college or post compulsory education institution and is where students can achieve the A-levels, Scottish Higher or other vocational qualifications (such as the former General National Vocation Qualification, for example) needed for university. In North America, however, community colleges are publicly or privately funded institutions of higher education which offer two-year courses of study typically leading to an associate’s degree. The associate’s degree is the lowest undergraduate degree in the hierarchy of post-secondary academic degrees offered in North America and is awarded upon completion of a course of study usually lasting two years. Common types of associate’s degrees are listed in Table 2.1. An associate’s degree is roughly equivalent to the Business and Technology Education Council's Higher National Certificate in Great Britain³.

² In North America the terms college and university can be used interchangeably.

³ This equivalency is based on the fact that the HNC and an associate’s degree are both worth 120 credits at Level 1 according to The Open University’s credit transfer system.

Table 2.1 Common Types of Associate’s Degrees (transfer degrees in bold)

Associate in Electronics Engineering Technology (AEET)	Associate of Fine Arts (AFA)
Associate in Engineering Technology (AET)	Associate of Forestry (AF)
Associate in Physical Therapy (ASPT-APT)	Associate of General Studies (AGS)
Associate of/in Arts (AA)	Associate of Industrial Technology (AIT)
Associate of/in Science (AS)	Associate of Occupational Studies (AOS)
Associate of Applied Arts (AAA)	Associate of Political Science (APS)
Associate of Applied Business (AAB)	Associate of Pre-Engineering (APE)
Associate of Applied Science (AAS)	Associate of Public Service (APS)
Associate of Arts and Sciences (AAS)	Associate of Science in Nursing (ASN), also
Associate of Arts in Teaching (AAT)	Associate of Nursing (AN) and Associate
Associate of Baccalaureate Studies (ABS)	Degree in Nursing (ADN)
Associate of Business Administration (ABA)	Associate of Software Development (ASD)
Associate of Engineering (AE), also	Associate of Technology (AT)
Associate of Engineering Science (AES)	

An associate's degree can be divided into two general categories: transfer degrees and career or professional degrees. The transfer degree forms the foundation of a bachelor's degree by allowing students to complete all of the general education requirements (see discussion below) prior to possible transfer to a four year college or university. [Transfer degrees include the Associate of Arts (usually awarded to students with a major in the social science or humanities); the Associate of Science (areas of concentration are usually in mathematics, natural sciences, health sciences, or technology); the Associate of Fine Arts (typically awarded to students in Music, Theater, Art, Dance, and Creative Writing); and the Associate of Arts in Teaching]. According to the National Center for Education Statistics (2010), 750,164 associate’s degrees were awarded in the US academic year 2007-2008, comprising over one-third of all undergraduate degrees awarded during that time. More than half of the associate's degrees were transfer degrees.

Students choose to attend a community college for a number of reasons, most notably finances (community colleges are often cheaper than four-year schools), grades (community

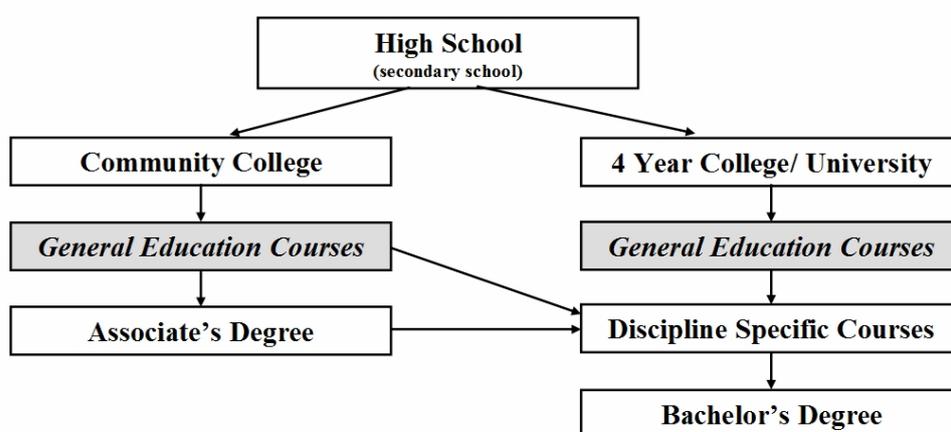
colleges have more lenient entrance requirements than four-year schools), or other personal reasons (students who do not feel ready to leave their parents' home, for example, often attend a community college in their area and continue to live with their parents). A community college student population often consists of persons who live in the local area, thus the name "community". The community college is a unique educational experience in many ways, and as such, research bodies exist for the professional development of instructors as well as the benefit of students; the American Association of Community Colleges has provided oversight on community college research since the 1920s, including publication of a research journal, *The Community College Journal*. Several peer-reviewed journals also extensively publish research on community colleges, such as *Community College Review* and *Journal of Applied Research in the Community College*. General education courses (discussed in section 2.2.3 below), including freshman composition, are an important aspect of the community college.

Students who enter a four-year college graduate with an undergraduate Bachelor's degree. To complete their course of study, students typically complete general education courses in the first two years, the freshman and sophomore years, then engage in discipline specific courses in the third and fourth years, the junior and senior years. Because general education courses are commonly similar among institutions of higher education, community colleges offer many of the same type of courses that four-year students take in the freshman and sophomore years (those typically taken to complete the associate's degree), but the community college does not offer discipline specific courses which juniors and seniors complete. Because community colleges do not offer discipline specific courses, many students who complete transfer associate's degrees (composed of general education courses) at a community college then undertake discipline specific courses at the four-year college or

university to earn the bachelor’s degree. Students attending a four year college take the same general classes their first two years, but usually do not receive an associate’s degree, as these students continue on to take discipline specific courses offered at the college, and then receive a bachelor’s degree (typically designed to be completed in four years of full time study).

Figure 2.1 illustrates the paths to a bachelor’s degree in North America. As in the community college, the freshman composition course is paramount in four year colleges.

Figure 2.1 Paths to a Bachelor’s Degree in North America



2.2.3 General education courses

North American undergraduate degrees are based on a credit system. Each course (or module) is allotted a specific number of credits (usually 1-4) depending on the number of class attendance and work hours required. Typically, classes are three credit hours, and approximately 130 credits are required for a bachelor’s degree; half of these credit hours are derived from general education courses. General education (GenEd) courses are those found in a “core curriculum” and are usually required for all students to graduate; GenEd courses are not program specific and are “intended to build domain knowledge but not expertise” (Geisler, 1994). See Appendix 1 for sample GenEd requirements. The GenEd courses that a student is required to take will come from disciplines outside of the student's major; given this, GenEd requirements may not be the same from school to school or even department to department, as

departments will often have their own set of requirements for students in their majors. Typically, however, GenEd courses are divided into sciences and humanities, including math, literature, writing, natural and social sciences, foreign languages, and fine arts and generally cover areas deemed to provide a person with a holistic, basic education. Freshman composition (discussed in-depth in section 2.2.4 below) is one such GenEd course. GenEd courses are designed not so students can develop strategic knowledge of the subject matter, but to deliver factual or declarative knowledge (Leki, 2007, p. 240). In sum, the GenEd program is intended to provide students with an introduction to a variety of disciplines, to ensure students are exposed to areas that facilitate a well rounded education.

Higher education courses, both GenEd and discipline-specific, have a common naming system across North America. They typically include a series of three to four letters followed by a series of three to four numbers. The letters are generated from the department which offers the course, ENGL for English, for example, and the numbers are generated for the level and credits of the course, 1013 or 2003, for a freshman year course and a sophomore year course, respectively, each worth three credit hours. Thus, the higher the course number, the more discipline-specific and difficult the course, e.g. HIST 414, a senior level course in the history department which only a student who has completed several HIST 100, 200, and 300 level courses would take. Only 100 and 200 level courses can be found in community colleges. Due to their basic content, 100 level courses are often taken by freshman students (those in their first year of higher education)⁴.

Many students do not choose a major directly upon entering higher education; in fact, nearly two-thirds of undergraduate students in the United States change majors before graduating and might consider up to four or five majors before finally deciding on one

⁴ This is why you often hear of something basic referred to as “101;” for example, “grammar 101” meaning grammar basics.

(Tobash, 2005). Since all students, regardless of major, take mostly GenEd courses the freshman and sophomore years, students who change their major or decide on a major late do not lose any time in completing their university requirements. Some controversy, however, is attached to the General Education curriculum requirement. On one end of the continuum, several smaller institutions have become famous for embracing a core curriculum that covers nearly the student's entire undergraduate education, often utilizing classic texts of the western canon to teach all subjects including science. St. John's College in Maryland, USA, is one such example. On the other end of the continuum are Ivy League schools such as Brown University in Rhode Island, USA, and Cornell University⁵, in New York, USA, who have largely done away with core requirements in their entirety, advocating for a student-driven course selection (Brown University, 2010). Although Leki (2007, p. 240) reports that "frequently interviewed faculty maintained that the main purpose of the introduction biology, geography, or political science [or other GenEd course] was to allow students to use the knowledge they developed in these courses in their daily lives," these applications have never been visited or tested. University students often lament GenEd course requirements, but institutions maintain their importance to developing a well rounded education [by helping students to think independently; understand and critically evaluate information; analyze and evaluate arguments; develop and present cogent written and oral arguments; explore one's own culture and history as well as those of others; understand, interpret, and evaluate the arts; and think critically about how individuals influence and are influenced by political, economic, cultural, and family institutions (University of Illinois at Chicago, 2008)].

⁵ Interestingly, the University of Birmingham (UK) was originally modeled on Cornell (The Carnegie Committee, Anon., 1899).

2.2.4 *Freshman Composition*

2.2.4.1 Background

Freshman Composition (freshman comp), recently re-named First Year Composition to avoid gender bias, is a series of GenEd courses designed to introduce students to the expectations of college writing and to help improve rhetorical skills to meet those expectations. The course began at Harvard College in 1885 in response to acceptance of a wider variety of incoming freshman; in addition to the typically elite students with boarding school and private educations, Harvard began accepting some middle-class students with less rigorous public educations. In order to ensure all students were ready and capable to engage in academic work throughout their university career, the English A course was designed to offer students intensive instruction and practice in oral and written communication which they could then carry forward to other classes. Today, though different schools often have different names for freshman comp courses (e.g. University Writing Program, Composition and Modern English, College Composition, College English), most are designed to meet the goals for successful completion set forth by the Council of Writing Program Administrators⁶ and include similar objectives: the study and practice of the process approach to writing in the academic community with a focus on critical thinking and research methods.

2.2.4.2 Organization

Though freshman comp courses are typically organized around writing types, such as persuasive, analytical or reflective writing, some courses are topically organized; for example, the Interpretation and Argument course at Carnegie Mellon University which focuses on the theme of *work*, with writing assignments such as writing a personal work biography and a

⁶ The Council of Writing Program Administrators is a national association of college and university faculty with professional responsibilities for (or interests in) directing writing programs. Members include directors of freshman composition, undergraduate writing, WAC/WID, and writing centers, as well as department chairs, division heads, deans, and so on (wpa.council.org).

three generation work history, or the Encountering Cultures course at Purdue University based on examining and responding to cultural issues. Even with such variation, freshman composition courses typically engage in the following assignment types (Wardle, 2009; Ferris, 2011; see also chapter 4):

- literacy or personal narratives, such as how one person in the home, school, or community shaped the student's development;
- rhetorical analysis of a text(s);
- persuasive writing, often based on information from two or more texts;
- research papers, self-selected or assigned topics which require the use of library, internet, and/or field research;
- analysis of visual texts (ads, art, etc);
- observation, such as observing events, people, or places in order to notice details and/or focus on a topic;
- a combination of these types, such as writing an argumentative research paper which compares and contrasts the types of knowledge claims made by two different genres of scientific writing that discuss a common topic.

Issue with these assignment types, and their tendency to be genres that do not respond to rhetorical situations requiring communication in order to accomplish a purpose (Wardle, 2009), is discussed below.

Typical topics in freshman comp include self (including self as a writer), popular culture, current events (such as the Japanese nuclear crisis), local issues (such as campus budget cuts), text analysis (written or visual), and, of course, teacher-selected themes (such as “cliques” or literary/film-based themes such as “vampires” or *Twilight*) (Ferris, 2011; see also chapter 4). Other key elements in freshman comp include audience awareness, thesis

formation and support, development or invention of ideas, organization and documentation, and understanding conventions in various disciplines. The specific freshman composition courses from which the texts used for the research reported in this thesis came are discussed in Chapter 4.

2.2.4.3 Controversies

As with the GenEd program in general, much controversy surrounds freshman comp. Firstly, according to Morris (2000) and Fulkerson (2005), the field of freshman comp is continually in a state of chaos. Many departments which house the course have an “anything goes” policy for overseeing it, and while many institutions have an optional or even required syllabus, there are generally no adopted required textbooks at large institutions; furthermore, few institutions or departments impose order on the course or incorporate substance into the course description. As Morris (2000) points out, as a practical matter, the sheer number of people teaching the course defies this imposition. To complicate matters, the writing discipline itself is engaged in “cultural” debates about what should be taught and learned in freshman comp (Durst, 1999), and there is a glaring lack of a strong consensus about what constitutes good freshman comp writing. Many of these debates revolve around the question of freshman comp as a “service course” (Morris, 2000), a course which teaches skills that can be directly applied in the study of other subjects. Should freshman comp be more concerned about teaching rhetoric as instructors understand it or should it be more concerned about training students to write papers that will earn high grades in other courses? Or, are those the same thing?

Regardless of this controversy, freshman comp is a staple of the university experience. According to Kroll and Alford (1997), in 1991, over 1.3 million students were enrolled in college writing classes in the community college alone. Moghtader, Cotch, and Hague (2001,

p. 457) found that the first year writing requirement is more widespread now than it was 25 years prior; in 1998, 97% of the more than 200 institutions surveyed reported having a writing requirement (as opposed to 76.5% in 1973); in fact, 100% of the public institutions surveyed reporting holding this requirement in 1998 (only 84% of public institutions required first year writing courses in 1973). Not only is freshman comp required in most of the approximately 4,000 colleges and universities in America, it is also often a prerequisite for upper level courses in other subject areas besides English. In this sense, freshman comp is generally regarded as a gatekeeper course; if students successfully pass freshman comp, it is assumed that they are able to produce college-level essays, reports, and answers. (Issues surrounding freshman comp as “gatekeeper” are discussed in more detail in section 2.4.6 below.)

The assumption that students who pass freshman comp are able to produce college-level writing presupposes the existence of a "universal educated discourse" (Russell, 1995) that can be transferred from one writing situation to another, and leads to a second controversy in the field of writing related to freshman comp: can one or two writing classes, in fact, prepare students for the writing they will encounter in their other university classes? At first glance, the answer seems promising. In a study investigating whether students were able to apply six freshman comp course proficiencies in the other writing situations they encountered in their university studies, Ahrenhoerster (2006) tentatively concluded that freshman comp did seem to positively affect the way students wrote in other courses, as papers written by students who had taken freshman comp performed noticeably better than those who had not on four of the course proficiencies: organization, developing ideas and constructing arguments, incorporating and documenting sources, and critically reading and thinking. While this does show transfer of the particular skill proficiencies assessed from

freshman comp, as Ahrenhoerster notes, we can not be certain that these same proficiencies are what faculty in other disciplines are actually looking for.

In fact, a number of composition researchers and theorists have, for over a decade, criticized freshman comp outright as a general writing skills instruction course that can not meet its objectives of preparing students to write in the university and beyond (see Gage, 1982; Freedman, 1995, Petraglia, 1995; Russell, 1995; Bartlett, 2003, among others). According to Bartlett (2003, p. A39), “professors cite a host of writing-related shortcomings among students, most often their inability to construct the sort of lengthy, sophisticated research papers required in upper-division courses.” Gage (1982, p. 469) reports that most faculty complain that “we can’t...see how what goes on in the typical freshman composition course can be of any use to us,” and he argues, along with Petraglia (1995), Russell (1995), and others, that freshman comp should be organized explicitly to teach the skills needed to read and write in the discourse of traditional academic disciplines.

However, as Downs and Wardle (2007, p. 552) write, “more than twenty years of research and theory have repeatedly demonstrated that such a unified academic discourse does not exist and have seriously questioned what students can and do transfer from one context to another (MacDonald, 1987; Ackerman, 1991; Carter, 1993; Kaufer & Young, 1993; Petraglia, 1995; Russell, 1995; Berkenkotter & Huckin, 1995; Diller & Oates, 2002).” Recent applied linguistics research involving writing in various disciplines also shows that it may not be possible to prepare students for the writing they will encounter in their discipline specific classes; while some general features of writing are shared across disciplines (Hyland, 2004; Swales, 1990), and there does exist a set of language features which characterize academic writing (discussed in Chapter 5), these shared features are realized differently within different academic disciplines, courses, and even assignments. For example, applied linguistics

dissertations use material and relational processes more frequently in the methodology sections (John, 2009) but scientific research articles use material processes in the method section and relational processes in the results and discussion section (Martinez, 2001). Humanities and social science papers use personal pronouns considerably more frequently than hard sciences and engineering papers; in the first, authors maintain a stronger identity in the research where authors in the second tend to highlight the issue(s) under study more than their personal role in the study (Hyland, 2002a). Disciplines also make variant use of introductory *it* plus that-clause and to-clause complementation; research articles in science (biology, chemistry, and environmental) use the structures significantly less frequently than research articles from business, language and linguistics, and public and social administration, but research articles from law use the structures more (Peacock, 2011). Social sciences employ stance features more frequently than natural sciences (Hyland, 1999; Charles, 2009). So in response to organizing freshman comp to teach the skills needed to read and write in the discourse of traditional academic disciplines, the question becomes “which academic discipline(s)?” As noted above, many students do not choose a major until their sophomore or junior year, and more than half of undergraduates change majors before graduating, perhaps considering up to four or five majors before finally deciding on one. Given this, combined with the unique attributes of individual disciplines, it would be impossible to identify exactly what type of writing the majority of students will engage in throughout their undergraduate studies, and therefore, which they should be taught in freshman comp.

An additional problem inherent in the idea of teaching academic writing skills in order to transfer writing knowledge to discipline specific courses stems from psychological views on transfer of learning. Based on his review of 100 years of research on transfer, Detterman (1993, p. 18) says that “if there is a general conclusion to be drawn from the research done on

transfer, is it that the lack of general transfer is pervasive and surprisingly consistent;” much of the research on writing transfer concurs with Detterman’s conclusion (Walvoord, 1985; McCarthy, 1987; Walvoord & McCarthy, 1990; Beaufort, 1999; James, 2010). Even Ahrenhoerster (2006), in his study of transfer mentioned above, was able only to evaluate the transfer of specific course proficiencies from freshman comp when evaluating students’ writing from other university courses; he was not able to ascertain whether those proficiencies that were transferred actually helped students score high marks on that discipline specific writing. Though far transfer—transfer of knowledge from one context (freshman comp) to another (biology lab reports)—is not likely, Ferris (2011) reported two key components that do appear to lead to successful transfer are reflection and learning of writing terminology such as “rhetorical situation,” “audience” or “genre.” As a solution to this problem of transfer of specific knowledge and the question of academic discipline, Downs and Wardle (2007) and Wardle (2009) purport the use of a first year composition pedagogy which seeks to improve students’ understanding of writing, rhetoric, language, and literacy in a course that is topically oriented to reading and writing as scholarly inquiry; in other words, “moving from teaching how to write in college to teaching *about writing*” (2007, p. 553) to help students understand that writing is content and context specific and, potentially, to transfer writing principles.

On the other hand, in her research on how students develop as writers, Carroll (2002) explains that although students claim to feel besieged by a barrage of disparate writing tasks in their first two years of college, tasks that might not be followed up in later years, students also report that they often return to what they learned in freshman comp classes (in part because the discipline-specific writing instruction they get in the third and fourth years is too specific). She maintains, though, that freshman comp can not meet all the needs of even more experienced writers, as “students’ complex literary skills develop slowly, often

idiosyncratically, over the course of their college years...” (Carroll, 2002, p. xi-xii). A study recently completed reported on at the 2009 BALEAP conference by Claire Furneaux and her colleague at the University of Reading likewise found that students require the entire three years of their undergraduate studies to become proficient writers.

Because some instructors and administrators struggle with the fact that freshman comp can not deliver the writing skills that students and faculty across the disciplines expect, they are increasingly under pressure to assess and demonstrate student learning in their courses. Yet, research on evaluation (Cooper & Odell, 1998), portfolios (Yancey & Weiser, 1997), grading (Allison, Bryant, & Hourigan, 1997), and assessing writing programs (Yancey & Huot, 1997), as well as any issue of the journal *Assessing Writing*, assert it is extremely difficult to identify what makes writing good (Carroll, 2002). One of the most useful definitions of good writing comes from Hjortshoj (2001, p. 33), who recognizes that

features of good writing vary from one situation to another. These variations depend, for example, on the *subject* of the writing, its *purpose*, and the *reader's expectations*. The *form* of writing used in a field of study often structures those expectations. As a consequence, the features of good writing in a literature course will differ greatly from the features of good writing in business or astronomy, and what seems clear to one audience might not be clear to another.

Students in Carroll’s (2002) study found, surprisingly to them, that learning how to figure out “what the teacher wants” was one of the best lessons they learned in their freshman comp course. This demonstrates the fact that, perhaps, they do understand good writing, as Lester et al (2003) found that 83% of students’ writing in all years of university study (freshman through senior), whatever the expressed function for that writing (report, persuade, analyze,

create a pattern of thought), was written primarily to receive a grade (as opposed, for example, to reveal their thinking in non-evaluative contexts).

Ultimately, based on the analysis of the data collected for her study, Carroll (2002) believes that although students in freshman comp may not necessarily learn how to write “better”, they do learn to write differently across the curriculum (in ways that may or may not be recognized by faculty), and the data collected in this study ultimately confirms that there is a limited but still useful role for freshman comp in the university.

2.2.4.4 Summary

The curricular, economic, political, and cultural issues embedded in freshman comp are too complex to be fully dealt with in this thesis, and because these debates have been going on since the foundation of English A at Harvard, they are not likely to be resolved in the near, or possibly far, future. Nevertheless, freshman comp is of importance to the university experience, and, therefore, to those ESL students who will encounter it upon completion of their English for academic purposes studies in an intensive English program.

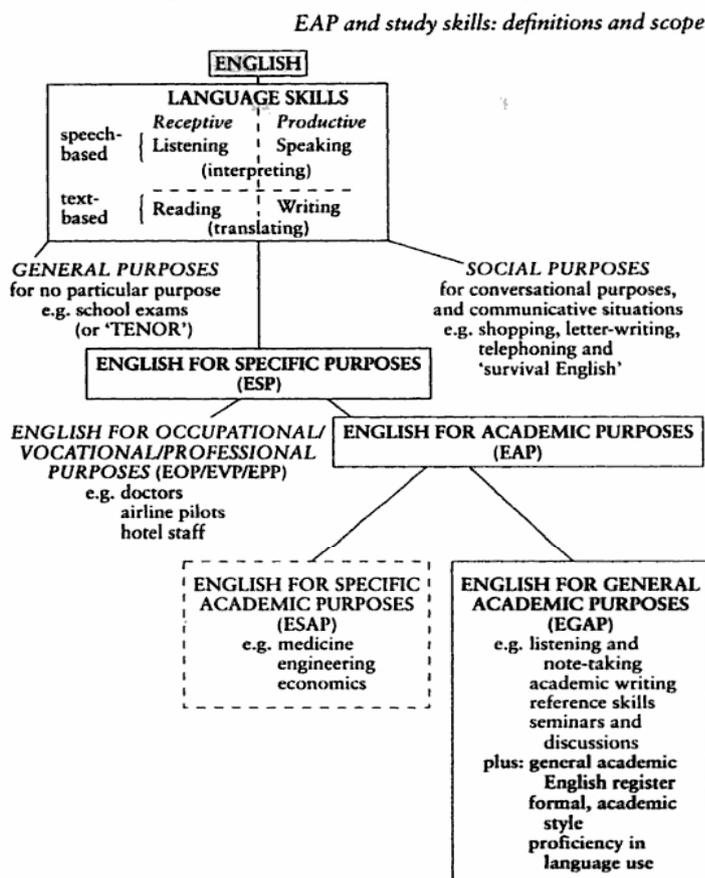
2.3 English for Academic Purposes (EAP) and the Intensive English Program (IEP)

2.3.1 EAP

The term EAP was first used in 1974 by Tim Johns at the University of Birmingham (Jordan, 1997, p. 1) and formally defined in 1975 as “concerned with those communication skills in English which are required for study purposes in formal education systems” (The British Council, 1975). By the end of the 1970s, the term EAP was in widespread use in both the UK and USA. As shown in Figure 2.2, EAP is a subset of English for Specific Purposes (ESP) and can be broken down into two further aspects: ESAP (English for Specific Academic Purposes) or EGAP (English for General Academic Purposes). ESAP is concerned with more discipline specific learning, such as that which takes place at the graduate level.

EGAP, as the name implies, is concerned with the language to be found in more general academic settings, such as the undergraduate setting where a variety of disciplines are introduced (in North America, specifically in the General Education courses discussed above.) For purposes of this thesis, EAP means EGAP.

Figure 2.2 EAP and study skills: Definition and scope. Source: Jordan, 1997, p. 3.



According to Flowerdew and Peacock (2001, p. 8) the context for teaching EAP generally occurs in four geographical domains: major English speaking countries (e.g. the UK and USA); former colonies of the first (e.g. Hong Kong, Singapore); countries with no historical links to the language (e.g. China, Norway); and countries of the former USSR (e.g. Ukraine, Kazakhstan). In these domains, students may need EAP for higher education studies in their own country (e.g. for reading academic texts) or for higher education in L1 countries (Jordan, 1997; Hyland 2006). In the North American context, it is typically students who

desire to earn a higher education degree from a college or university in that context—and whose first language is not English—that enroll in non-credit bearing EAP courses, in order to obtain English language skills which will allow them to successfully participate in academia. Broadly, those skills include taking on new roles, engaging with knowledge in new ways, and understanding different ways of constructing knowledge (Hewings, 2004; Hyland, 2006). Put simply, EAP practitioners “find out what the students have to do and help them to do it better” (Gillett, 1996, p. 1). As such, two indispensable characteristics of EAP are needs analysis and task-based learning/authentic learning activities (Gillett, 1996; Jordan, 1997; Flowerdew & Peacock, 2001; Ali & Yunus, 2004; Hyland, 2006).

EAP programs are “a practical affair...typically understood in terms of local contexts and the needs of particular students” (Hyland, 2006, p. 1) and often include teaching the four language skills (reading, writing, listening, speaking) and study skills (Jordan, 1997; Flowerdew & Peacock, 2001; Hyland, 2006). Invariably, it is the same three study skill areas that appear as the major causes of difficulty for students: listening and speaking seminars, academic writing, and listening and note-taking in lectures (Littlewood & Liu, 1996; Jordan, 1997; Hyland, 1997; Bhatia & Candlin, 2001; Evans & Green, 2007; Hüttner, 2008). Given these considerations, courses in an EAP program typically include those on academic writing, oral communication, and intensive reading. Of the courses offered in the EAP context, academic writing garners the most attention. Jordan (1997, p. 73-74), reporting on a survey conducted by BALEAP,⁷ described specific practices of EAP programs; not only was academic writing a course in every program, it was allotted nearly twice as much instruction time compared to other courses [e.g. (25%) for writing, (15%) for listening and note-taking, and (12%) for grammar]. Furthermore, academic writing has been identified as a central

⁷ British Association of Lecturers in English for Academic Purposes

concern for EAP students (Littlewood & Liu, 1996; Hyland, 1997; Bhatia & Candlin, 2001; Evans & Green, 2007) and has received the most attention in research related to EAP (Jordan, 2002). These views are evidenced in published textbooks for EAP courses, as well; of the more than 60 textbooks initially considered for a review of EAP materials, Tribble (2009, p. 400) identified nearly half as those which are used to teach writing. These facts are not especially surprising considering writing is a crucial element of assessment and occupies a gate-keeping status (discussed in section 2.4.5 below) in the university.

2.3.2 EAP writing

Given the high importance placed on writing, both in EAP programs and this thesis, it is important to consider approaches and materials used in teaching writing in the EAP. According to Hyland (2006, p. 1), EAP is “grounded in the social, cognitive and linguistic demands of academic target situations, providing focused instruction informed by an understanding of texts and the constraints of academic contexts.” This bears out in four research paradigms for investigating EAP writing (Flowerdew, 2002): genre analysis, contrastive rhetoric, ethnographic approaches, and corpus-based analysis. As is evidenced in chapter 3, genre and corpus-based analysis, and to a lesser extent contrastive studies, dominate the literature on L2 writing research. This may be due to the concern of EAP practitioners to understand “features of academic writing which could be passed on to students” (Hewings, 2004, p. 131-132). Following this aim, not surprisingly, many of the same issues plaguing freshman composition also permeate the teaching of EAP writing, in particular the debate regarding the effectiveness of teaching a general academic writing. Although the arguments presented in 2.2.4.3 stem from L1 writing researchers involved in freshman composition, the same arguments are presented by L2 writing researchers in regards to the EAP context: (1) Does a “general academic writing” exist? (2) Can academic writing

“skills” be taught with the expectation that they will transfer to writing completed in academic disciplines? Although some EAP practitioners advocate the teaching of general academic writing, expecting academic writing skills to transfer, current trends in teaching EAP writing answer no to both of the above proposed questions (Jordan, 1997; Flowerdew & Peacock, 2001; Hewings, 2006; Hyland, 2006). Generally, the issues relating to the teaching of academic writing in EAP are fundamentally those concerned with the field of L2 writing: comparison of L1 and L2 writers, the effectiveness of feedback and error correction; the role of grammar in L2 writing; and assessment. These issues are considered more in-depth in Chapter 3 (section 3.2.3).

In EAP, Tribble (2002, p. 131-132) says that teachers need to assist writers in three areas of responsibility: “(a) the processes that are necessary to the completion of a writing task, (b) the institutional and contextual constraints which operate in the target environment [to] determine what constitutes an allowable contribution, and (c) the linguistic choices which have to be made in order to produce such allowable contributions.” McKay (1994) sees the development of four aspects as crucial for fulfilling those responsibilities: 1. schemata (writing about content, not just “personal feelings”); 2. rhetorical patterns (becoming aware of students’ own tendencies, then noticing what happens in the texts they need to write); 3. social awareness (awareness of audience and conventions, understanding what should be said and how it should be said); 4. language (making appropriate choices, considering differences between spoken and written language, noticing what happens in the target texts). A main consideration for EAP writing materials includes a dual focus on both product and process. Tribble (2002) sees helping students recognize and understand the product, or kinds of texts students will ultimately need to compose, as a main aim of EAP writing; students “need to understand why texts are written in particular ways and what other texts they interrelate with,

and they need to be able to use the linguistic resources” characterizing those texts (Tribble, 2002, p. 145) in their own products. There is general agreement among EAP scholars that materials targeted at EAP writing should serve to understand product and fulfill the aim of “helping to socialise the student into the academic context” (Jordan, 1997, p. 166; see also, Flowerdew & Peacock, 2001; Hewings, 2004; Hyland, 2006). To assist in doing so, Tribble (2002, p. 146) suggests the use of exemplar corpora, corpora that are comprised of texts “which are the same as or, at least, analogous to the texts” that learners need to write. Exemplar corpora allow generalizations about the texts to be made; these generalizations can then be incorporated into instructional materials so that students can incorporate them into their own products. The NAFWiC, introduced and described in Chapter 4, is an exemplar corpus.

Furthermore, EAP writing materials should combine the focus on product with a process approach to writing (Jordan, 1997, p. 176). However, in the same study which identified nearly half of EAP pedagogical materials as those which are used to teach writing, Tribble (2009) categorized the overwhelming majority of those EAP writing materials as focusing singularly on process; these texts stem from an intellectual/rhetorical tradition, which is typified

by a focus on Exposition (including: exemplification, process, cause and effect, comparison and contrast, definition, division and classification), Description, Narration, Argumentation, and Classification, with writing tasks progressively moving from sentence, to paragraph, to whole text (Tribble 1996: 84–5). Methodologically, materials in this tradition have often been informed by Process Approaches to writing instruction and can be said to have a ‘focus on the writer’ (Raimes 1993). (Tribble, 2009, p. 404)

Thus, these materials focus on the process itself, often at the expense of the product. Tribble cites the *Longman Academic Writing Series* (which is discussed in chapter 7) as an example of materials rooted in the intellectual/rhetorical tradition. The topic of EAP writing materials is discussed more in-depth in Chapter 7, pedagogical applications.

2.3.3 *The IEP*

The Intensive English Program (IEP) is one particular English language teaching context in which EAP is often taught. Stoller (1994, p. 321-322) defined IEP as “post-secondary programs of instruction—in public or private tertiary institutions in the USA—designed to develop and strengthen the English language skills of persons whose native language is not English.” Typically, IEPs conform to a set of guidelines put forth by the National Association for Foreign Student Affairs and TESOL⁸. According to the American Association for Intensive English Programs, most IEP students are preparing to enter North American universities after completing an IEP course of study (AAIEP, n.d.). To further exemplify the role of EAP in the IEP, the University and College Intensive English Programs organization defines itself as an organization committed to, among other goals, educating institutions to the fact that adequate support of IEPs would lead to international students better prepared to succeed in their academic studies (UCIEP, 2011).

A typical IEP curriculum is designed for students at most levels of English language proficiency (absolute beginning levels are often not served in IEP programs), with an emphasis on developing the necessary oral and written skills for academic studies (or, on occasion, business or professional communication). The principles and teaching practices of EAP academic writing, discussed in section 2.3.2 above, are consistent with those in the IEP. Thus, the North American teaching context of the IEP is ideal for preparing students for the

⁸ Teachers of English to Speakers of Others Languages

North American freshman composition course. Applicable pedagogical implications are introduced below and further discussed in Chapter 7.

2.4 A Staircase Model to Inform EAP Writing Instruction in the IEP

2.4.1 Proposal of the model

As a pedagogical tool, MacBeth (2010) views models as not only basic, but also ubiquitous and invaluable. She comments, “It is hard to imagine a dental school without its big plastic tooth, health classes without a CPR dummy, or today’s higher tech computer simulations for anatomy” (MacBeth, 2010, p. 37). For novice writers—L1 or L2—who are unfamiliar with academic writing conventions, academic writing assignments are possibly confusing and frustrating (MacBeth, 2010, p. 34; see also Bizzell, 1982; Currie, 1998; Lillis, 1999); models may serve as “visible pedagogy” (Hyland, 2004, p. 8), reducing such unfamiliarity and its consequences. Though models are not without criticism [they control or inhibit students’ identities as writers, misrepresent the processes and styles of writing, and do not easily transfer to other writing tasks (MacBeth, 2010, p. 35)], they allow otherwise excluded participants an opportunity to become involved in discourse; first, by realizing the features and functions present in the model, followed by, likely gradual, independent construction of discourse leading to learning academic conventions through experience and participation.

For academic writing, as Biber, Gray, and Poonpon (2011, p. 16) note, an accurate description of the target register is needed; this need is two-fold: to determine what students need to know and to assess whether students are progressing toward that end. As Gillett (1996, p. 17) said, the EAP instructor’s job “is to find out what the students have to do and help them to do it better.” Discussed in section 2.2.4 above, all North American university-bound English for academic purposes (EAP) students, indeed all undergraduate students

regardless of first language, must complete first year undergraduate writing courses, despite the controversies surrounding them, in order to obtain the higher education degree they seek. Jordan (1997, p. 247) says that one of the best pieces of advice that can be given to students about to begin academic writing is to ask for good examples of appropriate writing; he further argues that EAP course directors should be able to obtain good specimens of writing to analyze and display, as in an exemplar corpus discussed above. “The primary focus should, thus, be on academic discourse genres and the range and nature of academic writing tasks, aimed at helping to socialize the student into the academic context” (Jordan, 1997, p. 166).

Numerous corpus-based studies on scholarly, and to a lesser extent graduate, writing (discussed in Chapter 3, section 3.5), have begun to contribute to a description of academic writing, and these lexicogrammatical findings are beginning to inform models of academic writing. The problem, however, is that research suggests a wide range of variation between novice and expert writers (Işık Taşa, 2010, p. 125). At the surface, this does not appear to present a problem: EAP writing pedagogy can highlight these differences in order to establish a benchmark for novice writers to work toward. Yet, descriptions of expert writing, or even those of graduate writing, may not be valid for undergraduate writing or freshman composition. In fact, studies have suggested that graduates and undergraduates have different needs in the nature of their writing requirements (Reid, 2001, p. 150; see also Hale, et al, 1996; Ginther & Grant, 1996; Hamp-Lyons & Kroll, 1997). For one, the purposes of student and expert writing differ; expert writing often presents new research findings and endeavours to convince readers of the importance of the research (Nesi, 2008), while student writing attempts to exhibit knowledge and learning (Hüttner, 2008, p. 153). Additionally, the use of expert writing as a standard for student writing is considered “both unfair and descriptively inadequate” (Lorenz, 1999, p. 14; see also Hyland & Milton, 1997). Furthermore (and

discussed in more detail in 2.4.3), modeling expert writing to novice writers constitutes far transfer, which is rarely effective (see also section 2.2.4.3). Perhaps most importantly, Bazerman (2006, p. 27-28) stresses that if we require novice writers to attempt to take on expert writing

prematurely, the authority of the disciplinary discourse may wash over and obliterate their ability as individuals to engage with and grow into disciplinary possibilities. We may put them into positions too distant from their current selves for them to make sense of.

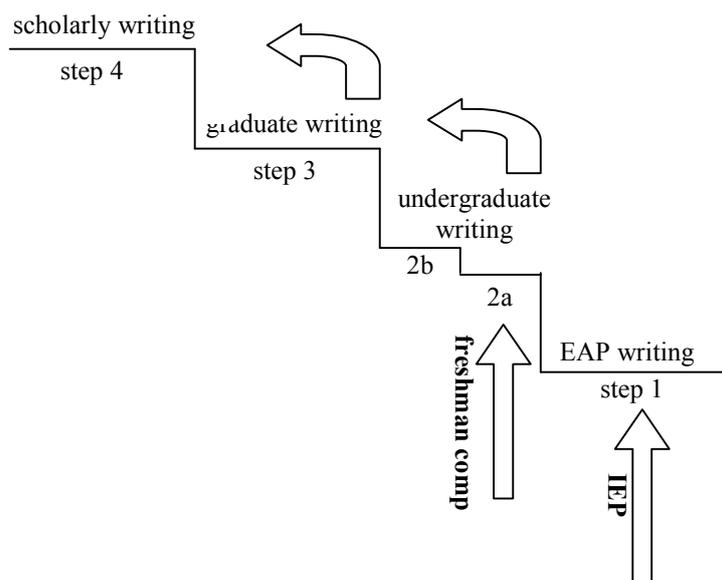
While novice writers are certainly capable of manifesting their own destiny, expert writing as a model is clearly not ideal for EAP writing.

Concerning the development of appropriate L2 writing models, Hüttner (2008, p. 147) maintains for the close analysis of student genres, especially those of undergraduates new to university. Similarly, as discussed in section 2.3.2 above, Tribble (2002, p. 146) discusses the use of exemplar corpora, comprised of texts which are the same as those that learners need to write. (The subject of which texts might be the same as those that EAP writers need to write is discussed in sections 2.4.3 and 2.4.4 below). However, use of student texts as exemplars may present other difficulties. Firstly, criticisms of student texts are strong, notably their apparent lack of communicative purposes (Hüttner, 2008, p. 147). Additionally, Hüttner (2008, p. 163) notes the decision of which texts constitute prototypical examples should be made by gatekeepers familiar with subject or disciplinary conventions. In response to the first difficulty with student writing as exemplar texts, acknowledging that EAP writers in the IEP will be required to engage in student writing in the university, thus classifying it as a need, recognizes “the legitimacy of considering pedagogical realities as *authentic* learning targets” (Hüttner, 2008, p. 149). The second difficulty of gatekeepers regarding student writing as

exemplar texts will be discussed in Chapter 4, section 4.3.3; suffice it to say here that this issue was adequately addressed in the exemplar corpus created for the research reported in this thesis.

To address the many issues with models discussed above, this thesis proposes a staircase model of the progression of academic writing skills in higher education in order to understand and inform effective writing instruction, or those academic discourse genres and range and nature of academic writing tasks alluded to by Jordan (1997). The model (see Figure 2.3) is composed of 4 steps in academic writing which could be labeled novice to expert, though here they have been identified by the context and setting in which they take place: scholarly, graduate, undergraduate, EAP. Materials stemming from the proposed model are achievable for students, reflect their communicative purposes, but are also acceptable to the gatekeepers of the discourse, thereby allowing learners “to engage with and grow into disciplinary possibilities” and become, as Hüttner (2008, 162) asserts, “*authors* rather than *copiers*.”

Figure 2.3 A staircase model of the progression of academic writing development in higher education



2.4.2 Description of the model

For the proposed model in Figure 2.3 above, Step 4, scholarly writing, is that writing which is written by experts in a subject for an audience with technical background on the subject; in other words, scholarly writing is published writing: composed by fully fledged experts on a topic read by other experts on the topic, representing the highest level of academic writing. Step 3, graduate writing, takes place in the context of graduate study; a natural progression from step 3 to step 4 takes place as graduate students are often refining their discipline-specific knowledge in order to become an active part of their discipline's academic community. Step 2, undergraduate writing, is broken into two substeps. As mentioned in sections 2.2.2-2.2.3, undergraduates in North American higher education spend the first half of their studies on general education courses; this is Step 2a. As undergraduates become introduced to their chosen discipline in the later years of undergraduate higher education and undertake discipline specific courses, they progress to substep Step 2b. A natural progression from Step 2b to Step 3 occurs at the onset of graduate study.

As noted in Figure 2.3, step 2a is where students who have graduated from high schools in North America often begin the development of their academic writing skills in higher education; this is done through the general education freshman composition course discussed above. While Step 4 is expert writing, Step 2 is novice writing. So what is step 1? The step below novice first year undergraduate writing is EAP writing; students who are studying to obtain the language skills necessary to begin step 2, undergraduate writing. As noted in Figure 2.3, step 1 is where EAP writing in the IEP takes place.

2.4.3 Theories informing the model

Constructivist thought emphasizes both the linguistic input and the social interaction involved in language learning. Vygotsky established one of the most crucial facets of social

constructivism, the zone of proximal development (ZPD). According to Vygotsky (1978), each learner has a ZPD, which is the distance between their existing developmental level and the level of potential development. In order for language learning to effectively take place, instruction must occur in the zone of proximal development. An important component of the ZPD, as it fits within social constructivism, is that the fundamentally human process of meaning making should occur in collaborative activity with other members of the culture. A number of applications of the ZPD have been made to language learning (e.g. Lantolf, 2000; Nassaji & Cumming, 2000; Marchenkova, 2005). Two principles of learning and the ZPD are key in the model proposed here: intersubjectivity and asymmetry (Vygotsky, 1978).

Learning must be intersubjective in that it involves a shared definition of situation, but asymmetric in that not all relationships in the situation are equal, with one relationship challenging another to develop. Importantly, Frawley (1997) notes that both intersubjectivity and asymmetry can be constructed and maintained by language; here, intersubjectivity manifests as academic writing and asymmetry as the steps on the model. As the process of joining the academic discourse community, climbing the staircase, is as much social as cognitive, the model in Figure 2.3 aims to target academic writing instruction in the IEP at the learner's ZPD. The ubiquity of the freshman composition experience in North American higher education is a social activity in terms of constructivism and fits within this framework.

The input hypothesis, one of the five somewhat controversial theories on second language acquisition purported by Krashen in the late 1970s and early 1980s, asserts that input is “the only true cause of second language acquisition” (Krashen, 1984, p. 61); for a learner to acquire language, input should be “‘a bit beyond’ his or her current level of competence.... If an acquirer is at stage or level i , the input he or she understands should contain $i + 1$ ” (Krashen, 1981, p. 100). In other words, learners should be able to understand

most of the language they are exposed to, but still be challenged to learn and understand some of it. As Brown (2007, p. 295) explains, “the corollary to this [$i + 1$] is that input should neither be so far beyond [the learners’] reach that they are overwhelmed ($\dots i+2$), nor so close to their current stage that they are not challenged at all ($i+0$).”

Among the criticisms of $i+1$ are its resemblance to Vygotsky’s ZPD discussed above; this is not a criticism according to this thesis, however, but rather a supporting factor. Another criticism of the $i+1$ theory is that Krashen presents the $i+1$ formula as if i and 1 may be defined, and as Gregg (1984) and White (1987), among others, have argued, this is not typically possible. In light of the factors at work in this thesis, however, while i and 1 are always precisely defined on an individual basis, generally speaking both the i and the 1 may be defined according to the steps on the staircase model for EAP writers in the IEP. As writers naturally move within their ZPD from undergraduate writing (step 2) to graduate writing (step 3), from graduate writing (step 3) to scholarly writing (step 4), progressing in the $i+1$ formula from step to step, the natural progression from the IEP (step 1) within the learner’s ZPD is to undergraduate writing; more specifically, freshman comp is step 2a. (The input formula $i+1$ also leads the learner from step 1 onto step 2.) In many contexts, freshman composition programs are the first place that students go after study in an IEP or other English language program, thus making freshman comp within the ZPD and $i+1$ of the IEP.

2.4.4 Freshman comp as a discipline

If a discipline can be characterized by the courses, assignments, and distinctive realization of shared features in its writing, evidenced in the discussion in 2.2.4.3, freshman composition may be seen as its own discipline. As Wardle (2009, p. 776) points out, and consistent with the findings in this thesis, particular assignments are common and recurring in freshman comp, “and thus, over time, they appear to have become genres,” and the linguistic

features in those genres is distinct from that of other genres. While this conclusion will, more than likely, be a hotly contested one (in accordance with other debates surrounding freshman comp), I believe further research examining writing which characterizes freshman composition will continue to support this conclusion.

It is interesting to note that Matsuda (2003) traces the role of L2 writing in North America back to issues with the teaching of freshman comp in the post-World War II induced growth of international students at tertiary institutions across North America. “With the continuing increase of international students in US higher education and the creation of disciplinary division of labor between L1 and L2 composition, preparing international ESL students *for required first-year composition courses* became an important responsibility for ESL teachers in intensive English programs” (Matsuda, 2003, p. 19, my emphasis). As such, the model proposed here brings the teaching of English for academic purposes writing in the intensive English program full circle. Although focus for ESL writing courses shifted to preparation for writing in discipline specific contexts, the view of freshman composition as a discipline as argued here reconciles the “old” approach with the “new” one.

2.4.5 Ascending the staircase

As Ferris (2011) notes, the transition for many L2 learners from intensive English program to freshman comp may often accompany various challenges.

In many contexts, the [freshman comp] classes are still primarily taught by inexperienced graduate students, who know little about teaching in general or composition teaching in particular and almost nothing about the characteristics of L2 writers. They often have little or no knowledge or preparation in linguistics and grammar, either, let alone understanding second language and literacy acquisition. Complicating the picture further is that composition instructors may often

misunderstand or even resent the needs of L2 writers in their classes. (Ferris, 2011, p. 14)

Undergraduate instructors often hold one of four attitudes towards the L2 writers in their classes (Ferris, Brown, Liu, & Stine, in press): it doesn't matter ("*It doesn't really matter whether these students are multilingual because they aren't having ESL difficulties. I don't need to know.*"); I warned them ("*I put at the top of my syllabus: 'This is not an ESL class.'*"); I feel their pain ("*I permit a few more errors...*"); and, it's complicated ("*Absolutely the first thing I do when I go to respond to a paper is to see who wrote it...to contextualize their needs...*"). In other words, L2 writers entering freshman comp may have different experiences based on the attitude of their writing instructors, but ultimately, the demands and expectations for students are high. To assist with the transition from EAP studies in the IEP (step 1) to undergraduate studies in freshman comp (step 2a), the following focuses have been recommended for IEP instruction (Zhu, 2004; Ferris, 2011):

- Strong reading skills;

"Many [freshman comp] classes demand a lot of reading—both teacher-assigned and student-selected for research projects—and often these readings are harder and longer than what [students] have encountered in ESL settings. If you teach at upper levels of an ESL program, you can help [students] a lot if you just push them harder in terms of reading expectations." (Ferris, 2011, p. 20)

- Preliminary research and citation skills;
- Awareness of plagiarism/textual borrowing rules;
- General sense of essay/paragraph organization. Freshman comp instructors will expect that students in their classes have been drilled in the 'five-paragraph essay'... They do not necessarily *like* the five-paragraph essay—on the contrary,

most...composition professionals despise it—but they at least assume that students have been taught about thesis statements, topic sentences, introductions, body, and conclusion, and transitions between paragraphs. They will not spend much, if any, time on this in a [freshman comp] class.... [so] we at least need to make sure that [students] know the general pieces of an academic essay. (Ferris, 2011, p. 20)

- Comfort working with peers;

Collaborative tasks are ubiquitous in composition today—and indeed in many disciplines and professions—and they go well beyond basic peer review tasks. Often students are asked to work in groups to write jointly authored documents, complete research projects, and give presentations. Part of the learning experience is figuring out how to understand and solve group problems....[if teachers have] stayed away from peer/group work because of cultural concerns, it is time to get over that, especially if many of your students want to continue their studies in English-medium universities. [Students] need preparation and practice in these situations. (Ferris, 2011, p. 21)

- Solid grammar and editing knowledge. As noted above, the demands and expectations for freshman comp students are high, and “faulty grammar and mechanics will be assessed harshly” in freshman comp, even though, given “the aforementioned issue of instructors having no training in grammar (let alone grammar pedagogy),” students will receive very little, if any, in-class instruction or assistance. Therefore, “L2 students need to have a solid foundation in sentence-level grammar” (Ferris, 2011, p. 20). The research in this thesis addresses this last recommendation of solid grammar.

One problem, however, as will be discussed in some depth in Chapter 3, is that research in academic writing, particularly corpus-based research, has focused on either

features of scholarly writing (Steps 3 or 4) or learner writing (Step 1), with little focus on undergraduate writing (Step 2). In order to facilitate a natural progression from step 1 to step 2, undergraduate writing, particularly writing from step 2a, first year undergraduate writing, must be researched so that instructors can be aware of the grammar used in successful freshman comp so they may ensure students understand it.

2.4.6 Potential criticisms of the model

Although Vygotsky and constructivist ideas inform some of the ideas presented here, one criticism which could be aimed at the proposed staircase model and teaching implications of this thesis could stem from critical pedagogy, specifically relating to teaching to freshman comp. Critical pedagogy, primarily concerned with critiquing existing educational institutions and practices with aims to transform both education and society (Freire, 1972; Giroux, 1981; Apple, 1982), sees freshman comp as part of the very existing power structure which impedes language learners' individual voice, particularly for those who may believe that multilingual students should simply integrate into the academic mainstream unquestioningly. However, critical pedagogy also seeks to involve learners in the understanding, application, and production of knowledge as it matters to their lives, situations, and needs. As Bizzel (1982, p. 196) asserts, "what they need from their education is...the critical training to trace their victimage to social forces rather than to 'fate,' and hence to work toward control of their own destinies." In this sense, preparing students to successfully ascend the staircase to the undergraduate step 2a and join the academic discourse community there, as basic or controversial as it may be, prepares language learners to attain this sort of critical understanding while also giving them the opportunity to continue their climb up the staircase.

Additionally, criticism of this model could stem from the idea of freshman comp as a “gatekeeper” course, mentioned in section 2.2.4.3 above. As a gatekeeper at the university, freshman comp may, in some ways, determine who will succeed to higher levels of education, and therefore, possibly higher levels of affluence; here, freshman comp as gatekeeper is a reflection of and contributor to unequal power relations. However, this type of extreme linguistic pluralism may actually work against language learners by not providing students with the linguistic tools they need to enter the English-speaking mainstream (Flowerdew, 2002; Baker, 2008). Equipping students to participate in the discourse community can, as Baker (2008, p. 248) maintains, also embrace “an inclusion of the students’ own rich and valid culture backgrounds” while at the same time facilitating their ascension of the staircase.

Another potential criticism of the staircase model and associated ideas advanced in this thesis regards upholding the “native speaker standard.” It should be noted that the staircase and its implications in no instance refer to “native” or “non native” speaker status, but rather focus on context of discourse; in fact, no distinction has been made regarding native or non native speaker status in the freshman comp essays which comprise the corpus in this study. As Silva (1997, p. 216) notes, “a credible general theory of writing must be based on more than research on the writing of native English speakers,” and Leki (1992) found, concerning compositing processes, at least, that “the distinction to be made was not between L1 and L2 writers but between experienced and inexperienced writers” (in Hamp-Lyons & Kroll, 1997, p. 3), which is the distinction made in the staircase model. In this case, the research focuses on writing deemed to be successful in the chosen context, aligned from novice to expert, regardless of speaker status; essays of both first and second language English users were included in the corpus, as both users demonstrated successful writing in the context.

One last potential criticism of the staircase model and ideas presented here may relate to the seemingly linear depiction of the progression of academic writing skills. In some linear models, such as the one presented here in Figure 2.3, it is assumed that learners must master one stage before going to the next, and returning to an earlier stage is considered regressing. I view the model here, though, more in line with Ausubel's subsumption theory (Ausubel, 1963, as cited in Brown, 2007, p. 84), which describes meaningful learning as a process of relating and understanding new material to relevant, existing concepts; as learners advance each step of the staircase in meaningful context, the new language and skills they encounter and acquire interact with and are subsumed under a more inclusive system, confirming their meaningfulness.

2.4.7 Applicability of the model

One final word in this chapter concerns the wider applicability of the research presented in this thesis. Although the proposed model and research in this thesis is directly applicable to EAP students studying in an IEP with aspirations to undertake North American undergraduate studies, every academic context is different. In some countries—in fact, probably most outside of North America—there are no freshman composition programs, and, therefore, the implications of the research as discussed in chapter 7 may not be directly applicable in these contexts. Nevertheless, the model proposed in Figure 2.3 above holds true for any EAP context: students who are studying in pre-sessional or other courses at level 1 with aims to engage in undergraduate study should engage in materials which help advance them to step 2 ($i+1$ in their ZPD), not step 4 or even 3.

2.5 Summary

To establish the context of the research presented in this thesis, Chapter 2 discussed the institutional frameworks of the North American higher education context, the concept of

general education courses and the important aspects of the freshman composition course as part of that context, as well as explored teaching EAP in the IEP. Additionally, the relevance of freshman composition to that of L2 writing in IEPs was established, and a staircase model was proposed as a basis to inform academic writing instruction in the IEP.

Freshman composition is a ubiquitous, though controversial and contentious, course in the North American higher education context. Specialists within the freshman comp field often find themselves questioning the purpose of the course, and therefore, how the course should be taught. Regardless of the controversies and negative connotations surrounding freshman comp, it is a salient feature of the university experience for all undergraduates and typically engages students in assignments such as rhetorical analysis, research papers, and literacy narratives. As the staircase model for the progression of academic writing skills in higher education proposed in this chapter illustrates, freshman comp is the next step for the English language learner who wishes to embark on an undergraduate education in North America; as such, using the freshman comp course as a focus for academic writing instruction in the IEP can efficiently prepare language learners for a successful climb up the academic writing staircase.

CHAPTER THREE

RESEARCH IN SECOND LANGUAGE WRITING AND CORPUS LINGUISTICS

3.1 Introduction

This thesis argues that materials and methods for English for academic purposes (EAP) writing instruction in the intensive English program (IEP) should be heavily influenced by freshman composition as it is taught in North American higher education contexts. As previously discussed, three areas of research are relevant to the arguments presented here: the North American higher education context, specifically the course of freshman composition; the state of research on second language (L2) writing in the English as a second language field, specifically that relating to corpus-based studies of academic writing; and the teaching of L2 writing in EAP, specifically in the IEP. Chapter 2 explored freshman composition as part of North American higher education and the teaching of EAP in the IEP in order to establish the research context for this study and propose a model informing the research here; this chapter reviews selected works in the areas of corpus-based studies of L2 writing that are relevant to this study. Specifically, an overview of L2 writing research is provided in section 3.2, and section 3.3 presents a synopsis of research stemming from corpus linguistics. Section 3.4 establishes the relationship between corpus-based studies and L2 writing research, with section 3.5 focusing specifically on research conducted in the contexts appearing on the staircase model of progression in academic writing skills in higher education introduced in Chapter 2—scholarly writing, graduate writing, upper-level undergraduate writing, lower-level undergraduate writing, and learner writing—to establish the gap in the literature addressed by this thesis.

3.2 Second Language (L2) Writing Research

3.2.1 History of L2 writing

While research in second language writing has a complex history, struggling to establish disciplinary and organizational affiliations (Leki, 2000), an understanding of the discipline of second language writing studies situates itself “between concerns of compositionists and applied linguists” (Kroll, 2003, p. 12), as it considers students not only as novice writers but also as second language learners. Though there was little published research on L2 writing 25 years ago, the researching and teaching of L2 writing has, in the past decade, steadily become more prevalent, with researchers from English language teaching, applied linguistics, communication, composition studies, and education identifying themselves as L2 writing specialists. Atkinson (2000) considers the field to be primarily North American based, and this is evidenced by the entities through which L2 writing has established itself as a discipline. In 1998, CCCC (Conference on College Composition and Communication) formed the Committee on Second Language Writing to integrate the second language perspective into institutional practices of CCCC. The Symposium on Second Language Writing (SSLW) also began in 1998 as a biennial conference to bring together teachers and researchers who work with second- and foreign-language writers to discuss important issues in the field of L2 writing (due to overwhelming popularity and success, SSLW became an annual international conference in 2006). In 2005, the members of TESOL voted to begin a new interest section devoted solely to issues in L2 writing, and more and more TESOL graduate programs are offering L2 writing courses, whether as required or elective. This is not to say, however, that L2 writing research is conducted only in North American contexts; as is illustrated in section 3.5 below, strands of L2 writing research do focus on writers in foreign language settings.

3.2.2 Theory and methodology in L2 writing

Though L2 writing has seemingly established itself as a viable discipline, and the past 25 years have seen an “exponential” growth of research in L2 writing (Polio, 2003, p. 35; see also Matsuda, et al, 2003), “...a single, comprehensive theory of L2 writing is perhaps a long way off...” (Ferris & Hedgecock, 2005, p. 3). This may be because, as Hinkel (2011, p. 535) notes, “the range of settings and contexts where L2 writing is taught and learned is enormous, as are the types of learners who set out to attain language proficiency and skills requisite to produce quality L2 writing.” The lack of a comprehensive theory of L2 writing may also stem from the variant influence on L2 writing from other key disciplines, namely rhetoric and composition, applied linguistics, and TESOL (Raimes, 1991, 1998; Grabe & Kaplan, 1996, 1997; Matsuda, 1998, 1999, 2003a, 2003b; Leki, 2000; Hedgecock, in press as cited in Ferris & Hedgecock, 2005, p. 3). While Hinkel (2011) claims that rhetoric has had a minimal influence on the investigations of L2 writing, most scholars believe that discussions of L2 writing research should begin with a discussion of contributions from L1 writing theory (Hamp-Lyons & Kroll, 1997; Kroll, 2003; Matsuda, 2003; Silva & Leki, 2004; Ferris & Hedgecock, 2005; di Gennaro, 2006); these are, broadly, the idea of distinguishing between novice and expert writers, the role of schema in accessing and creating written text, the illustration of cognitive processes involved in producing a piece of writing, and the examination of similarities and differences between L1 and L2 writing (Johns, 1995; Ferris & Hedgecock, 2005; di Gennaro, 2006). Though not explicitly focusing on differences *between* novice and expert writers, the research reported here does examine expert writing in order to influence novice writing instruction, thereby creating the argument that this study is influenced by L1 writing theory.

Further trends in L2 writing methodology have, of course, developed. The first of these maintained a focus on form and production, engaging writers in controlled compositions to practice lexical and grammatical forms (Silva, 1990; Matsuda, 1999; Kroll, 2001; Ferris & Hedgecock, 2005). A second trend in L2 writing methodology maintains a focus on the writer engaged in the composition process, encouraging readers to focus on fluency and ideas (Raimes, 1991; Ferris & Hedgecock, 2005; di Gennaro, 2006). In response to these trends, two additional L2 writing teaching methods developed congruently; these methods focus on disciplinary content and practices and communities, emphasizing learners' need to write for academic audiences within established genres (Ferris & Hedgecock, 2005, p. 13). Based on these pedagogies, writing instruction centers on noticing, practicing, and producing the features of texts written for particular audiences, as this study does with EAP writers and freshman composition. An additional L2 writing methodology recently developed maintains a focus on sociocultural issues and critical pedagogy, emphasizing that L2 writing instruction can not be "neutral, value-free, and nonexclusionary" (Belcher & Braine, 1995, p. xiii), but that "sociopolitical issues affecting life in and outside of academic settings" must be also be examined (Benesch, 2001, p. xv); these views were discussed as a potential criticism of the research and implications of this thesis in Chapter 2. While I agree with critical pedagogues that these ideas are not value free, the aim is to assist L2 writers to enter this otherwise "exclusionary" academic discourse and succeed in freshman composition, thereby appealing to "Freirean notions of liberatory literacy practices" (Grabe & Kaplan, 1996, p. 32).

3.2.3 Research trends in L2 writing

In characterizing L2 writing research, Archibald and Jeffrey (2000, p.7) highlight the fact that "classroom practice and observation has often been the source" for L2 writing studies, with pedagogical implications a focus of the research (Hinkel, 2011). In addition to this

overarching focus, surveys of L2 writing research have established paradigms which drive that research; these paradigms have been classified into similar but not identical categories (Silva, 1990; Raimes, 1991; Hyland, 2002b; Polio, 2003; Ferris & Hedgecock, 2005), though they generally fit within four types: writers, writers' texts, readers, and contexts of writing. It is obvious that there is potential for overlap between these categories. Researching writers and writers' processes involves looking at decisions and practices of writers, often to understand the influence of contextual factors on those processes; researching writers' texts and contexts of writing mainly results in an understanding of how writers use language in specific contexts to successfully communicate with readers, often to understand what makes a piece of writing successful by discovering what is typical in that type of writing. This thesis falls into the latter two categories, examining writers' texts in contexts of writing, researching successful writing in the specific context of freshman composition.

Additionally, nine topics of inquiry have consistently been explored in L2 writing research over the last several decades (Panofsky, et al, 2005; Jayne, et al, 2011; Hinkel, 2011):

- 1) theories of writing pedagogy, such as Hubert and Bonzo's (2010) study on the impact of L2 writing research in North American foreign language instruction or Raimes's (1991) earlier look at (then) emerging traditions in methods for teaching writing;
- 2) practices of writing instruction and professionalization, such as Rhee's (2010) investigation of Korean graduate students' textual borrowing, Casanave's (2010) case study of three doctoral students writing qualitative dissertations, or Ferris and Hedgecock's book presenting pedagogical approaches to teaching L2 writing (now in its second edition);

- 3) the role of grammar and/or vocabulary in writing, such as the use of collocations and lexical bundles (Chen & Baker, 2010; Taiwo, 2010), cohesive devices (Hinkel, 2001, 2002), or Norris and Ortega's (2000) meta-analysis of over 40 studies looking at explicitly taught grammar-focused instruction versus implicit instruction;
- 4) feedback on writing, such as the somewhat controversial and on-going debate on the effectiveness of teacher feedback to improve accuracy (Truscott, 1996; Ferris, 1999; Truscott, 1999; Ferris, 2004; Truscott, 2010; Ferris, 2010; Suh, 2010) or investigating learners' use and understanding of peer feedback (Zhao, 2010);
- 5) collaborative aspects of writing, such as effectiveness of peer response groups (Mittan, 1989; Liang, 2010), students' reaction to the use of peer response groups (Leki, 1990; Schmid, 1999), or peer collaboration in composing like L. Lee's (2010) case study of wiki-mediated writing;
- 6) technology in writing, such as earlier studies examining the effectiveness of teaching L2 writing in a computer lab (Bernhardt, Edwards, & Wojahn, 1989; Bernhardt, Wojahn, & Edwards, 1990) to recent studies on the role of the internet in the writing process (Warschauer, 2010; Conroy, 2010; Radia & Stapleton, 2010);
- 7) identity in writing, such as critical pedagogy in L2 writing instruction (Canagarajah, 2002) or academic biliteracy challenges and pedagogies (Cho, 2010; Butvilofsky, 2010);
- 8) genre and/or corpus studies of writing, from Swales's (1990) groundbreaking work presenting a methodology for genre analysis or Biber's (1988) initial work on textual dimensions of speech and writing to more recent and finely focused studies such as those on argumentative essay writing (Ong & Zhang, 2010; Qin & Karabacak, 2010; Ismali, 2010), as well as investigations on discourse frameworks and text properties of

L2 writers (Hinkel, 2003, 2004a, 2004b) and error types in L2 texts (Schleppegrell, 2002); and

- 9) assessment of writing, such as investigating the use of rubrics (Haswell, 1998; Reza Rezaei & Lovorn, 2010) or automated scoring systems (James, 2006; Enright & Quinlan, 2010; Dikli, 2010) in assessing L2 writing.

Again, the potential for obvious overlap is evident among each of these categories. For example, Bitchener & Knoch's (2010) "Raising the linguistic accuracy of advanced L2 writers with written corrective feedback" deals with grammar, feedback, and technology in writing. An additional topic covered in the research, though not distinguished in the categories above, mainly due to overlap, is researching the writing of academic disciplines, such as Zhang and Xhang (2010), "Thematic progression and discourse coherence of college English writing;" Bacha (2010), "Teaching the academic argument in a university EFL environment;" and Mungra and Webber (2010), "Peer review process in medical research publications: Language and content comments" (categorized as grammar and vocabulary, genre studies, and professionalization, respectively). As mentioned in Chapter 2, section 2.3.2, these topics also drive much of the research aimed to influence EAP practices.

Research on L2 writing addresses a wide range of issues in a number of fields; this is not surprising given the inter-disciplinary nature of L2 writing mentioned in section 3.2.1. One of the main areas of research in L2 writing centers on a genre-based approach. Again, this is not surprising given the current methodological approach of emphasizing learners' need to write for academic audiences within established genres as discussed in 3.2.2 above. But corpus-based research on L2 writing is also becoming more prevalent. The interface of corpus linguistics and L2 writing, and to a lesser extent genre, are discussed more in section 3.4. Because of their relevance to the research in this thesis, selected L2 writing corpus

studies are examined in section 3.5 below. What we see there are studies which focus on three areas of L2 writing; two of these areas, the role of grammar/vocabulary in writing and the research of writing in academic disciplines, is expected, given the association of grammar and vocabulary with corpus linguistics and the question of discipline-specific writing explored in Chapter 2. The third area, however, is a bit more unexpected: identity. While the focus on identity is obviously relevant for L2 writing studies, the idea of approaching identity using corpus-based methods was intriguing.

3.3 Research in Corpus Linguistics

3.3.1 History of corpus linguistics

In addition to the general state of research on L2 writing in the English as a second language field, research in the field of corpus linguistics is also relevant to this study.

Similar to that of L2 writing, the value of corpus linguistics as a discipline has expanded rapidly in the field of English language teaching in the past two decades, with specialists from a variety of second language and related disciplines using corpora and corpus methods to inform language study, and, more recently, language instruction. Like the *Journal of Second Language Writing*, the *International Journal of Corpus Linguistics* is less than 20 years old; in 1994, the TALC (Teaching and Language Corpora) Conference became one of the first international conferences with a focus on the role of corpora in language teaching. In North America, The American Association of Corpus Linguistics was formed in 1999; 2001 saw the first international Corpus Linguistics Conference, and in 2003, the British Association for Applied Linguistics established the Corpus Linguistics Special Interest Group. Schlitz (2010) purports that corpus-based approaches to describing and analyzing language have actually dominated linguistic studies for nearly 50 years, and Hunston (2002, p. 1) asserts that “it is no exaggeration to say that corpora, and the study of corpora, have revolutionised the

study of language, and of the applications of language, over the last few decades.” While corpora may have permeated L2 research in Europe since the latter half of the 20th century, its application in North American research and contexts has taken much longer (Simpson & Swales, 2001; Barlow, 2011), only recently establishing itself as a viable method of mainstream research. Nevertheless, “the contribution of corpus linguistics...is difficult to dispute” (O’Keeffe, McCarthy, & Carter, 2007, p. 21), as it is making outstanding contributions to the fields of second language research and teaching.

3.3.2 Theory and methodology in corpus linguistics

“Studies of language can be divided into two main areas: studies of structure and studies of use” (Biber, Conrad, & Reppen, 1998, p. 1). Studies of structure look at what is theoretically possible in a language; studies of use look at how language is actually used. Corpus linguistics engages in studies of language use by investigating patterns which occur in authentic language of various contexts. The use of corpora makes these patterns, which would more than likely otherwise go unnoticed through introspection alone, apparent (Biber, Conrad, & Reppen, 1998; Biber & Conrad, 2001; Hunston, 2002; Römer & Wulff, 2010; Hüttner, 2010). However, “it is accepted that corpus linguistics is not a theory-rich field” (Barlow, 2011, p. 4; see also Stubbs, 2006; Hunston, 2006; Gries, 2010). While Barlow (2011, p. 5) acknowledges this could be due to several factors—corpus linguists may be more drawn to data than notions of theory or may be content to use corpus techniques for their own purposes without being overly concerned about theory—he attributes the lack of theory in corpus linguistics more likely to the rejection of established theoretical constructions about language. Varying views derive from this notion, leading to a lack of consensus about the status of corpus linguistics: Is it a methodology? Theory? Discipline? All of the above?

Some scholars argue that corpus linguistics began as a methodology but has developed into a discipline with its own research agenda that goes beyond methodology (Aarts, 2002; Römer & Wulff, 2010; Tribble, 2010). Evidence of this transformation are various theoretical concepts and frameworks that have emerged from corpus-linguistic approaches to language, such as Hunston and Francis's (2000) *Pattern Grammar*, Hoey's (2005) concept of *Lexical Priming* (Römer & Wulff, 2010), or the integration of data, description, theory, and methodology (Laviosa, 2002) in corpus studies. Other scholars, however, believe that corpus linguistics is only a methodology (McEnery & Wilson, 2001; McEnery, Xiao, & Tono, 2006; Hardi & McEnery, 2010), or both a method and a discipline, with a methodology that can be applied to other disciplines or its own (Meyer, 2002, Flowerdew, 2002; Bowker & Pearson, 2002; Mukherjee, 2005, 2010; Wynne, 2010; Williams, 2010). This latter belief is evidenced by studies in other disciplines such as cognitive linguistics (Schönefeld, 1999), construction grammar (Goldberg, 2006), discourse analysis (Sotillo & Starace-Nastasi, 1999; Fairclough, 2000; Baker, 2006), psycholinguistics (Tomasello, 2003; Ellis & Larsen-Freeman, 2009), sociolinguistics (Holmes & Sigley, 2002; Friginal, 2009; Baker, 2010), and contrastive studies (Laviosa, 2002; Granger, Lerot, & Petch-Tyson, 2003; Butler, Gómez-González, & Doval Suárez, 2005; Johansson, 2007; Gomez-Gonzalez, Lachlan Mackenzie, & González Álvarez, 2008), among others, which have followed a corpus-based approach (discussed below). Yet others consider corpus linguistics a theory (Leech, 1992; Stubbs, 1993; Tognini-Bonelli, 2001; Teubert, 2005), a new paradigm from which language can be described, where "theoretical standpoints arise from practical investigations and not the reverse" (Williams, 2010, p. 401). This latter stance is the most controversial, with many corpus linguists seeing corpus-driven approaches as little more than an ideal. As McEnery, Xio, and Tono (2006) note, corpus linguistics "certainly has a theoretical status. Yet theoretical status is not theory

in itself.” Furthermore, “no method of working is neutral with regard to theory” (Hunston, 2002, p. 92; see also Halliday, 2005; Gries, 2010; Barlow, 2011); thus, it is not possible to approach a practical investigation to language without a theory already in place. That being said, a theory which originally informed a study can be modified based on the findings from that study. This line of thought leads to a “strong” and “weak” view of corpus linguistics as a theory. The strong view, little more than ideal, subscribes to the idea that “theoretical standpoints arise from practical investigations and not the reverse;” the weak view, more realistic and realized in such works as *Pattern Grammar* (Hunston & Francis, 2000), maintains that “corpus linguistics should lead to revised theories of language” (Hunston, 2011).

Corpus linguistics is clearly a discipline, and section 3.3.3 reviews the research trends in the discipline. Corpus linguistics is also clearly a method, as evidenced by the list above of studies in other disciplines conducted using a corpus-based approach. The distinction, however, is not always clear cut. For example, the study reported in this thesis uses a corpus-based approach to investigate frequency and lexicogrammar to contribute to a description of register, key elements of research trends of the corpus linguistics discipline, thereby adding support to corpus linguistics as a discipline; yet the applications of the findings to inform L2 writing pedagogy also center it within the L2 writing discipline, meaning that corpus linguistics is only a methodology employed in another discipline¹. Like L2 writing, corpus linguistics is highly inter-disciplinary. While the inter-disciplinary feature of L2 writing likely stems from the exploitation of writing from a variety of approaches, the inter-disciplinary feature of corpus linguistics likely stems from the exploitation of a highly productive method. As discussed in relation to pedagogical implications in Chapter 7, there is

¹ The issue of discipline distinction, including the evaluation of this thesis, will be discussed in section 3.5.2.

not necessarily a right or wrong answer to discipline or method, but varying applications and approaches for certain contexts.

Regardless of position as theory, method, or discipline, corpus analysis, in addition to being approached from a use perspective, is distinguished by four major characteristics (Biber, Conrad, & Reppen, 1998, p. 4). Corpus-based studies are empirical (analyzing the actual patterns of language use in natural texts); utilize a large and principled collection of natural texts as the basis for analysis; make extensive use of computers for analysis; and depend on both quantitative and qualitative analytical techniques. Furthermore, two approaches characterize studies in corpus linguistics (Biber, Conrad & Reppen, 1998; Hunston, 2002). Biber, Conrad, and Reppen (1998, p. 6) simply refer to the first approach as “investigating the use of a linguistic feature.” Hunston uses the term “word-based,” which I apply here. Word-based studies focus on the investigation of words, identifying meanings, patterns, and associations of that word; word-based studies typically begin at the sentence level and make use of concordance lines to access language in a corpus. The second approach for studies in corpus linguistics is what Biber, Conrad, and Reppen (1998, p. 6) simply refer to as “investigating varieties or texts.” Hunston uses the term “category-based.” I believe the term “context-based” is more descriptive. Context-based studies typically begin with the investigation of a register (or a context within a register), identifying the combination and application of words and categories which distinguish that context from another, often making use of statistical calculations and corpus annotations. This thesis engages in a context-based method of corpus linguistics, making use of frequency lists, tagging, and parsing to investigate how the noun phrase behaves differently in scholarly writing and first year undergraduate writing. As with the status of corpus linguistics, however, there is not necessarily, nor should there be, a clear divide between word- and context-based approaches.

As Hunston (2002, p. 94) says, word-based and category-based (or context-based) methods of corpus analysis can inform one another, much as qualitative and quantitative methods of research complement each other.

3.3.3 Research trends in corpus linguistics

Though, as discussed above, the corpus approach has been used in nearly all branches of linguistics, including, for example, lexicography, grammar, language variation, contrastive and translation studies, diachronic studies, semantics, pragmatics, stylistics, sociolinguistics, discourse analysis, forensic linguistics, and language pedagogy (Hunston, 2002; McEnery, Xiao, & Tono, 2006; Römer & Wulff, 2010; McEnery & Xiao, 2011), this section explores research trends in the discipline of corpus linguistics, though this is also not without debate. Thompson and Hunston (2006, p. 3) suggest there is no common set of research questions for corpus linguistics, while Biber, Conrad and Reppen (1998) see the discipline, broadly, serving to answer two fundamental research questions: what particular patterns are associated with lexical or grammatical features? How do these patterns differ within varieties and registers? And Barlow (2011, p. 3) identifies “three areas in which corpus linguistics has made a significant contribution to our understanding of language” use: phraseology, frequency, and register description. Publications in corpus linguistics have also led to applications in language teaching. General trends in the literature for these four areas of corpus linguistics are discussed below.

1. Phraseology is a central feature of corpus linguistics, rooted in Sinclair’s (1991) discovery that meaning is derived not from individual words but through phrases, several words in a sequence. Phraseology typically includes the study of collocations and lexicogrammar, including lexical bundles. An idea which is basic to the discussion of corpus-based phraseology studies is Sinclair’s (1991) open-choice and idiom principles. Simply put,

the idiom principle states that receivers of the English language do not understand a grammatical pattern with lexical words filling that pattern, but that understanding of meaning arises from the whole phrase rather than the individual words in it. Hunston (2002, p. 143) puts it in these terms: “if we look at English from the point of view of the words that make it up, then, each word can be described in terms of its preferred phraseologies.” The open-choice principle is less predictable, “with word-choice constrained only by the general grammatical rules of English” (Hunston, 2002, p. 145). The idiom and open-choice principles are important to our discussion of corpus-based phraseology because, as Hunston (2002, p. 147, 149) notes, “the idiom and the open-choice principle together provide a theoretical account for two observations: that phraseology is extremely pervasive in English and that phraseology alone can not account for how sentences or utterances are made up.”

The most prominent way of studying phrases is through collocation, which most corpus linguists agree is “the statistical tendency of words to co-occur” (Hunston, 2002, p. 12; Greenbaum, 1974; Sinclair, 1991; Hoey, 1991; Stubbs, 1995; McEnery & Wilson, 2001). Stubbs (2001, p. 58) says that “words are typically used in routine phrases, and that even the most frequent words have typical collocates and typical uses;” he goes further to say that “words are not chosen freely, but co-selected with other words in a span of a few words to left and right” (Stubbs, 2001, p. 84). Studies of collocation reveal a tendency for each collocate of a word to be associated with a single sense of that word; for example, when paired with collocates *good/great, deal* means “amount;” however, when paired with collocate *big, deal* conveys a lack of importance (Biber, Conrad, & Reppen, 1998). Studies of collocations can also uncover different uses of nearly synonymous words; for example, Kennedy (1991, p. 107) found that *between* is typically used after nouns like *differences, distinction, agreement, and meeting*, whereas *through* is more frequently found after verbs such as *go, pass, run, and*

fall. Semantic prosody (Louw, 2000) is also an aspect of collocation. Semantic prosody expresses a speaker's underlying attitude, much like the connotation of a word; for example, *fan the flame* (Stubbs, 2001b) and *bordering on* (Schmitt & Carter, 2004) typically carry a negative understanding where *provide* (Stubbs, 1995) and *career* (Stubbs, 2001b) typically carry a positive understanding. Other studies on collocation include applications for teaching English for specific purposes, including business (Walker, 2011), engineering (Ward, 2007), and general English for academic purposes (Durrant, 2009).

Lexicogrammar is another aspect of phraseology. Lexicogrammar is also rooted in Sinclair (1991) and his idea that there is no difference between lexis and grammar, or that lexis and grammar are so closely intertwined that they can not be productively studied separately². The *Collins COBUILD English Grammar* (Sinclair, et al, 1990) was the first work to make extensive use of this principle. One example given in the grammar is that “words with similar behaviours tend to have similar meanings. For example, nouns followed by *for* mostly indicate a reaction or feeling towards someone or something” (Hunston, 2002, p. 104). I would argue that the *Longman Grammar of Spoken and Written English* (Biber, et al, 1999), which plays a central role in this thesis (discussed in chapters 4 and 5), clearly makes use of the principle of lexicogrammar; in fact, features are examined in the *Longman Grammar* according to register distribution (discussed below), discourse factors, and lexicogrammatical patterns. The lexicogrammatical principle is realized through discussions of grammatical structures (e.g. verb tense, stance, complement clauses) accompanying lists of most frequently employed lexical items [*bleed, chase, shop, and starve* in the progressive tense in conversation (Biber, et al, 1999, p. 471); *actually, really, and probably* as stance adverbials in conversation (Biber, et al, 1999, p. 982); *think* and *say* as verbs taking *that-*

² While it has been noted that the term “lexicogrammar” itself was initially used by Halliday, Sinclair’s notion of the inseparability of lexis and grammar is typically represented, especially in the North American context, with this term and so it is used here.

clauses in post-predicate positions (Biber, et al, 1999, p. 662)] to achieve a communicative purpose. Other works investigating aspects of lexicogrammar include *Pattern Grammar* (Hunston & Francis, 2000), which describes words and their patterns and those associated meanings, such as the adjective *afraid* where the pattern “*afraid* + that clause” is associated with the meaning of *afraid* as apologizing for something; Hoey’s (2005) lexical priming, which argues that as language users encounter and use words in spoken and written discourse, they automatically pick up their usage patterns and learn in which language structures, textual positions, or text types the words typically appear; and Römer and Schulze’s (2009) collection of case studies providing evidence for the inseparability of lexis and grammar, such as Duguid’s (2009) research on how creative metaphors exploit grammatical patterning and semantic preference to achieve pragmatic effects.

Another area of phraseology that has generated research in corpus linguistics, especially in the North American context, is lexical bundles. Lexical bundles are “extended collocations: bundles of words that show a statistical tendency to co-occur...sequences of word forms that commonly go together in natural discourse” (Biber, et al, 1999, pp. 989-990), a sort of combination of collocation and lexicogrammar. Most of the work on lexical bundles has been done by Doug Biber and his colleagues at Northern Arizona University, USA. Perhaps the most comprehensive study on lexical bundles includes the work in the *Longman Grammar of Spoken and Written English* (Biber, et al, 1999) mentioned above. Here, Biber et al (1999) focus on lexical bundles which are three or more words in length, occur in the registers of conversation and academic prose, and appear at a minimum frequency; for example, the cut off for four-word bundles was at least ten times per million words. One key point in the findings on lexical bundles is that the bundles are not “expressions that speakers would recognize as idioms or other fixed lexical expressions” (Biber, et al, 1999, p. 990), yet

they still conform to patterns. Although “Sinclair suggests that any group or sequence of words is constructed and understood in the light of one or the other of the principles, but not both” (Hunston, 2002, p. 145), some lexical bundles could be viewed as a both- and category for the idiom and open-choice principles. While meaning is attached to the set patterns (idiom principle), what follows the bundles is only constrained by the general grammatical rules of English (open-choice principle). For example, the most common type of four word lexical bundles in conversation include the pattern “subject pronoun followed by a verb phrase, and, in many cases, the verb phrase extended by the beginning of a following complement clause” (Biber et al, 1999, p. 1002): *I + know—I don’t know what+, well I don’t know, I don’t know how+, I don’t know about+*. The idiom principle is realized by the use of the recurring phrases to mark meaning of personal stance; the open-choice principle is realized by the fact that the words following the bundles are open to choice, constrained only by the general grammatical rules of English. What it is that I don’t know (as in the bundle “I don’t know what”) will be grammatically constrained by the complement clause, but it could be an infinite number of possibilities: *I don’t know what—she had for supper; her name is; is going on; we should do; the result will be; this is going to mean politically; it is; the guy’s got to do; happens; you’d do with it; the joke is; the numbers are; it’s called; they’re doing; to tell you; that lady keeps talking about; time we go; else; a trickle is; his divorce settlement is costing; you think I am; being black has to do with it, etc.*³ Other studies on lexical bundles include investigating the use of lexical bundles across disciplines (Cortes, 2004, 2006; Hyland, 2008; Pecorari, 2009), registers (Biber, Conrad, & Cortes, 2004; Nesi & Basturkmen, 2006; Biber & Barbieri, 2007; Cortes & Csomay, 2007, 2009; Chen & Baker, 2010), as well as English as a lingua franca settings (Jablonkai, 2010). Lexical bundles is actually only one, of

³ Examples taken from the Corpus of Contemporary American English (corpus.byu.edu/coca).

several, terms used to describe “extended collocations;” other terms include phrasal lexemes (Moon, 1998), prefabricated patterns (Granger, 1998), chains (Stubbs & Barth, 2003), and chunks (Khuwaileh, 1999). These investigations also typically investigate language across registers.

Studies on phraseology are the most frequent of those conducted on scholarly writing discussed in section 3.5.2 below.

2. Frequency information has made a significant contribution to our understanding of language use. Frequency is the simplest kind of information available from corpus analysis; as Barlow (2011, p. 7) notes, “what computers do well and do easily is count, which means that whatever categories are marked in a corpus—words, lemmas, part-of-speech tags, etc.—tend to be counted.” As discussed above, the idea of identifying patterns in language is a key principle of corpus linguistics, and frequency information is one method for determining those patterns; as Hüttner (2010, p. 200) observes, frequency data aids researchers in identifying “typical patterns of language use” that “frequently escape intuitions of native speakers and of teachers.” The first, and perhaps most influential, research involving frequency was in lexicography, with the application of frequency information to dictionaries, where the most frequent meaning of a word can be listed first and less frequent meanings, perhaps, excluded altogether, or new meanings which had not been previously accounted for as being relevant can be included (Hunston, 2002).

Though frequency may be dismissed as theoretically irrelevant, it plays an important role in grammatical explanations (Barlow, 2011, p. 8). For one, frequency information provides “a focus on *typical* forms of expression rather than on the range of *possible* forms of expression” (Barlow, 2011, p. 7). The *Cambridge Grammar of English*, another work which contributes to this thesis and discussed in chapters 4 and 5, presents probabilistic rules of

grammar, meaning that the rules presented “state what is most likely or least likely to apply in particular circumstances” (Carter & McCarthy, 2006, p. 5). According to Carter and McCarthy (2006), while it is not necessary for learners to mimic native speakers’ use of language to be successful, it is important for learners to observe and understand how and why speakers use the language they do; describing language in use is not a prescription for learner use, but a presentation of data so that teachers and learners can make their own informed choices, a significant aspect of language learning (Carter & McCarthy, 2006, p. 10). This idea of making informed choices is discussed again in Chapter 7 in relation to the pedagogical implications of this thesis.

Additionally, frequency highlights what Scott (2000) refers to as keyness, which concerns the understanding of texts and content through key word forms, lemmas, or strings (Bondi & Scott, 2010); frequency has played a role in a wide range of studies concerning keyness (Tribble, 1999; Kemppanen, 2004; Scott & Tribble, 2006; Rayson, 2008), providing potential lexical insight into a discourse community. Frequency studies are also useful for guiding materials writers on determining what linguistic features to include in textbooks and other pedagogical materials (Hunston, 2002; Biber & Reppen, 2002). Three areas in which materials writers should consider frequency information include (1) grammatical features to include or exclude; (2) the order of grammar topics; (3) specific words to include when illustrating a grammatical feature. The ideas specifically relating to the first area, grammatical features to include or exclude in teaching materials, will be revisited in chapter 7 in the discussion of pedagogical applications of the research in this thesis.

As with the topics of inquiry discussed in relation to L2 writing research in section 3.2.3 above, there is overlap in the research of corpus linguistics, as well. Studies on

phraseology—lexical bundles and collocations—center on frequency, may be relevant to language teaching, and contribute to descriptions of register, as discussed below.

3. An ever-growing number of publications in corpus linguistics have also led to applications in language teaching. These publications generally fall into three categories: overview of major language teaching applications, overview of the field as a whole (including language teaching applications), and practical means for realizing those applications in the classroom. Publications in the first of these categories offer an overview of major developments in corpus-based language pedagogy, perhaps citing examples of how corpora have been used in the classroom or discussing its potential uses there (e.g. Conrad, 2000; Biber & Conrad, 2001; Tognini-Bonelli, 2001; Biber & Reppen, 2002; Sinclair, 2004; Gabrielatos, 2005; Galloway, 2005; Gavioli, 2005; Braun, Kohn, & Mukherjee, 2006; Campoy, Gea-Valor, & Belles-Fortuno, 2010; Harris & Moreno Jaén, 2010). Publications in the second category seek to provide an overall introduction to corpus linguistics and corpus-based methods, namely to students of applied linguistics (e.g. Biber, Conrad, & Reppen, 1998; Kennedy, 1998; Hunston, 2002; McEnery, Xiao, & Tono, 2006; Adolphs, 2006; O’Keeffe, McCarthy & Carter, 2007; Teubert & Cermakova, 2007); these books typically begin by introducing key concepts such as corpora, concordancing, collocation, lexicogrammar, frequency, and annotation, followed by a discussion of current research applying corpus-based methods in the field of corpus linguistics. The third category of publications in corpus linguistics that have led to applications in language teaching include those which endeavor to actually instruct teachers in how to bring corpora into the classroom (e.g. Flowerdew, 1996; Cobb, 1997; Conrad, 1999; Gavioli, 2001; Tribble, 2001; Hadley, 2002; Rob, 2003; Cheng et al., 2003; Coniam, 2004; Gaskell & Cobb, 2004; Lee & Swales, 2005; Tsui, 2005; Bennett, 2010; Reppen, 2010). These publications generally offer one or all of three alternatives for

classroom-based corpus linguistics: adopting corpus informed textbooks, using personally developed teaching materials for data-driven learning, and engaging students in activities with online corpora.

While a few of the publications discussed above are articles, the overwhelming majority are full length books; and while those books often provide examples of corpus-based studies, such as those discussed in section 3.5 below, to illustrate key aspects of corpus linguistics, they themselves do not typically report original research, and therefore, are not included in section 3.5 below. One notable exception is Biber, Conrad, and Reppen (1998), part of the second category of publications, who conduct a multi-dimensional analysis of research articles in ecology and history (included in Table 3.1 below).

4. The final area to be discussed in which corpus linguistics has made a significant contribution to our understanding of language use is register description. Studies which contribute to a fuller understanding of register description often fall into one (or both) of the areas described above (phraseology and frequency), and may use multi-dimensional analysis (Biber, 1988) (Charles, Pecorari, & Hunston, 2009). Studies contributing to a description of writing (Moon, 1998; Hyland, 1998, 2002c, 2004, 2008; Biber & Conrad, 1999; Stubbs & Barth, 2003; Biber, Conrad, & Cortes, 2004; Cortes, 2004, 2006; Biber & Barbieri, 2007) and speaking (Biber & Conrad, 1999; Khuwaileh, 1999; Biber, et al, 2002; Nesselhauf & Römer, 2007; Biber & Barbieri, 2007) include those on lexicogrammar and lexical bundles. Perhaps the most significant work contributing to register description is the previously mentioned *Longman Grammar of Spoken and Written English* (Biber, et al, 1999); in this reference grammar, a full description of both the structure and use of grammatical features of English is provided along four main registers: conversation, news, academic writing, and fiction. As

mentioned above, the findings presented in the *Longman Grammar* concerning the register of academic writing (discussed in chapter 5) are a noteworthy part of this thesis.

Furthermore, the development of register-specific freely available corpora, and publications stemming from them, are also leading to fuller descriptions of academic registers, namely the Michigan Corpus of Academic Spoken English (e.g. Yoo, 2008; Lee, 2009; Belles-Fortuño & Campoy-Cubillo, 2010), Michigan Corpus of Upper-level Student Papers (e.g. Ädel & Garretson, 2006; Wulff & Römer, 2009; Römer & Wulff, 2010), British Academic Spoken English (e.g. Nesi & Basturkmen, 2006; Low, Littlemore, & Koester, 2008; Lin, 2010), and British Academic Written English (e.g. Gardner, 2008; Bruce, 2010; Durrant & Mathews-Aydmh, 2011). Because of their relevance to the research reported in this thesis, selected studies focusing on descriptions of the written academic register are discussed in section 3.5 below.

Biber, Conrad and Reppen (1998) see studies in corpus linguistics investigating what particular patterns are associated with lexical or grammatical features and how those patterns differ within varieties and registers. The research reported in this thesis addresses both of these aspects, identifying noun phrase modification patterns and their frequencies in a corpus of freshman composition. Of the four areas in which corpus linguistics has made a significant contribution to our understanding of language use—phraseology, frequency, register description, and language teaching—the research here directly addresses the last two. By investigating differences between scholarly and novice writing the findings reported in this thesis contribute to the overall understanding of academic writing. Chapter 7, which discusses pedagogical implications of the motivations driving this research addresses applications for language teaching. As such, this thesis makes an original contribution to the research in corpus linguistics.

3.4 Corpus-Based L2 Writing Research

3.4.1 The need for a corpus-based approach

Although not all L1 writers, indeed probably few, maintain an absolute control of linguistic features and functions of language, the vast majority possess sufficient control for grammatical ability not to be a major focus of analysis in, or instruction for, L1 composition. Perhaps not surprisingly, one of the most noticeable differences between L1 and L2 writing is the role of linguistic competence; “based on a vast body of research, limited vocabulary and grammar are the most frequently cited/noted properties of L2 text” (Hinkel, 2011, p. 529). Additionally, L1 writers may have some understood knowledge of cultural expectations for writing, whereas L2 writers may not only have differing knowledge of expectations for writing, but, as noted above, are often still developing proficiency in the L2, making grammatical form as demanding as content (di Gennaro, 2006). Thus, while L1 composition provided a foundation for the writing component of L2 writing research, it was unable to adequately address these and other factors related to the “second language” nature of L2 writing; for this, L2 writing researchers needed to turn to their other “parent” discipline (Silva & Leki, 2004), that of applied linguistics (di Gennaro, 2006, p. 6). As discussed in section 3.2.3, genre- and corpus-based research in L2 writing is becoming prevalent. As writing is a multidimensional form of communication that involves control of rhetorical, linguistic, and social conventions (Matsuda 1998), the use of these approaches, especially a corpus-based approach is a logical choice.

While L1 writing researchers began to focus on process, L2 writing research shifted to a focus on linguistic features of student texts (see, e.g., Gipps & Ewen, 1974; Cooper, 1976; Flahive & Snow, 1980; Ferris & Politzer, 1981 as cited in Biber, Gray, & Poonpon, 2011, p. 6). This trend has continued so that it is common to find L2 writing research focusing on

measures of fluency, accuracy and complexity (see e.g., Ellis & Yuan, 2004; Brown, Iwashita, & McNamara, 2005; Larsen-Freeman, 2006; Nelson & Van Meter, 2007 as cited in Biber, Gray, & Poonpon, 2011, pp. 6-7) or applying corpus-based methods

...in a range of areas, including vocabulary knowledge (e.g. Nation, 2001), genre knowledge (Tribble, 2001; Tribble, 2002), grammatical knowledge (e.g. Clear, 2000; Diniz & Moran, 2005), ...citation practices (e.g. Thompson & Tribble, 2001)...as well [as]... both ...L1 and...[L2] writing research (e.g. Henry & Roseberry, 2001). (Schlitz, 2010, pp. 92-93)

Furthermore, several recent publications have highlighted the intersection between L2 writing and corpus linguistics. Articles in a special issue of *The Journal of Writing Research* (2010) used corpora to investigate topics relating to L2 writing such as writing development, writing assessment, and writing instruction (Schlitz, 2010, p. 93). Similarly, articles in a special issue of *IEEE Transactions on Professional Communication* (2006) were devoted to answering the question of what corpus linguistics could contribute to research in professional communication (Römer & Wulff, 2010). Additionally, three collections explore the interface between discourse/genre analysis—a traditional approach to L2 writing research—and the corpus-approach (Biber, Connor, & Upton, 2007; Ädel & Reppen, 2008; Charles, Pecorari, & Hunston, 2009). Charles, Pecorari, and Hunston (2009, p. 4) attribute the concern of “what Sinclair and Coulthard (1975) called ‘latent patterning’—the recurrence across many texts of detailed features of expression that play a crucial, though often hidden, role in the construction of academic knowledge” to the attraction between L2 writing research and corpus linguistics. Moreover, as discussed above, the need to focus on factors related to the “second language” nature of L2 writing, particularly linguistic features and functions, has led to this convergence. As is typical for L2 writing, nonetheless, the integration of L2 writing

research and corpus-based methods is driven “...by the desire to investigate features of academic writing...[and] to bring those discoveries to the classroom...” (Charles, Pecorari, and Hunston, 2009, p. 6).

The corpus-based research of L2 writing is of central importance to this thesis, specifically research which describes writing in a variety of academic settings. As such, selected corpus-based L2 writing studies focusing on descriptions of the written academic register are discussed in section 3.5.

3.4.2 Genre-based approaches

Before continuing this review of relevant research, it is prudent to discuss genre-based pedagogy here, if only in brief, due to its prevalence in the literature in L2 writing research (see sections 3.2.3 above and 3.5.4 below), and its similarity to corpus-based methods. Genre-based pedagogy is a popular framework for investigating the form and function of typically written texts in academic settings. Genre scholars propose that genre-based studies can help non-native speakers of English master the functions and linguistic conventions of texts that they will encounter in the academic setting (Swales, 1990; Bhatia, 1993; Flowerdew, 1993). Genre scholarship has been most fully developed in three research traditions (Hyon, 1996; Flowerdew, 2002): English for specific purposes (ESP), North American new rhetoric, and Australian systemic functional linguistics (SFL). New rhetoric genre studies tend to focus on purposes of texts, giving little attention to form, and utilize ethnographic analysis, while SFL and ESP genre studies often employ linguistic analysis emphasizing structural forms and features, or moves, characteristic of different texts. SFL and ESP studies are typically differentiated by the contexts in which they operate, namely primary and secondary school and non-professional workplace writing for SFL and university and professional writing for ESP (Hyon, 1996; Flowerdew, 2002). Thus, the investigation reported in this

thesis might be classified as ESP genre scholarship given its focus on linguistic analysis of form in a socially constructed university context to help English learners master the linguistic conventions that they will encounter in their undergraduate studies. We will return to this classification below.

There is some overlap in genre- and corpus-based studies. Hüttner (2010) describes her investigation of student paper conclusions as a corpus-based genre analysis; Flowerdew (2008) uses keyword analysis to characterize move structures in a genre. Hyland (1998, 2001, 2004, 2005, 2008) has done considerable work using a corpus approach to investigate genre, including the use of corpora and frequency reporting. Although there is potential overlap in genre- and corpus-based methods, there are important methodological distinctions between the two approaches. Firstly, genre-based studies are not concerned with investigating a large collection of texts in a corpus; as few as two or four prototypical texts are deemed sufficient. Additionally, genre-based studies rely more on qualitative, manual methods of analysis. Finally, genre-based studies do not typically investigate specific lexicogrammatical features but concentrate more on rhetorical moves (Flowerdew, 2002). This last distinction is particularly important in considering the literature reviewed in section 3.5. Each of these distinctions is also relevant to the study here; while it might be classified as ESP genre scholarship given its focus on linguistic analysis of form in a socially constructed university context, this research is singularly corpus-based because of the mixed qualitative and quantitative methods of a bottom-up, micro-level approach focusing on recurring lexicogrammatical patterns in a sufficiently large corpus⁴ (i.e. no top-down, macro-level approach investigating moves and structures has been applied, thereby rejecting any classification as genre-based). Interestingly, when proposing this study during one of the first meetings with

⁴ The issue of size for a corpus is discussed in chapter 4.

my supervisors in the summer of 2008, a question addressing this precise issue was posed: “Why corpus-based and not genre-based?” The simple answer at that time was, as discussed above, that the need to focus on factors related to the “second language” nature of L2 writing, particularly linguistic features and functions, can be best accomplished through a corpus-based approach; and, as Barlow (2011, p. 5) surmised in relation to the theory debate in corpus linguistics, I am more drawn to data stemming from lexicogrammatical analysis. As will be discussed in relation to further research in Chapter 8, however, a combined corpus- and genre-based study using data from the research reported in this thesis could be extremely productive for the pedagogical model proposed in Chapter 2.

3.5 Research Represented in the Levels of the Staircase Model

3.5.1 Organization

As detailed above, developments in both corpus linguistics and L2 writing research have led to extensive corpus-based research focusing on the academic written register. Selected studies conducted in the contexts appearing on the staircase model of progression in academic writing skills in higher education introduced in Chapter 2—scholarly writing, graduate writing, upper-level undergraduate writing, first year undergraduate writing, and learner writing—are reviewed here in order to establish the gap in the research addressed by this thesis.

3.5.2 Scholarly writing (step 4)

The overwhelming majority of corpus-based writing research has focused on the level of scholarly writing, that which is written by authors with a graduate or postgraduate education to readers of a technical audience and which has been published; scholarly writing is expert writing represented on level 4 of the staircase model. Perhaps the largest contribution to a description of scholarly writing as a whole comes from *The Longman*

Grammar of Spoken and Written English (Biber, et al, 1999) mentioned several times in the preceding discussions. It is from this reference grammar that the linguistic features and functions of scholarly writing which form a backbone of this thesis (discussed in chapter 5) were extrapolated. The findings presented in *The Longman Grammar* are of a general academic prose, as 13 disciplines comprise the corpus from which the findings were taken; as discussed in Chapter 2 (as well as in section 3.5.7 below), the notion of a general academic writing, even at the scholarly level, is debatable. Perhaps for this reason, many of the studies in L2 writing corpus-based research which focus on scholarly writing are typically discipline specific. Additionally, some studies focusing on explicit genres have also attracted attention. An overview of the research in scholarly writing is provided in Table 3.1, as well as discussed below. It should be noted that only research which specifies all categories in the table (discipline, genre, grammatical feature) is included here; for example, Biber's (2009) study on multi-word patterns in research articles and academic textbooks is not included because the disciplines comprising the corpus were not reported in the study. Furthermore, although some varieties of genres are presented in Table 3.1, the number is surprisingly few, especially compared to the number of disciplines investigated. This could be attributed, in part, to the fact that the "research article" is the most prevalent genre of scholarly writing; this could also be attributed to the current status quo of separation between genre and corpus-based studies (as mentioned in section 3.4.2), with genre studies focusing on moves rather than grammatical patterns. For example, Golebiowski's (2009) investigation of applied linguistics and education abstracts examines only the rhetorical structure of the abstract without reporting on any lexicogrammatical findings; therefore, it is not included here. This collection is by no means exhaustive, but attempts to represent a selection of available research.

Table 3.1 Selected Studies on Scholarly Writing (Step 4)

<i>Author/Year</i>	<i>Discipline</i>	<i>Genre</i>	<i>Grammar feature</i>
Butler 1990	Science	Research article	Modals
Gledhill 1996	Medicine	Cancer research	<i>In</i>
Williams 1996	Medicine	Research report	Lexical verbs
Jabbour 1997	Medicine	Research article	Tense
Biber, Conrad, & Reppen 1998	Ecology, history	Research article	Multi-dimensional analysis
Williams 1998	Biology	Research article	Collocations
Hyland 1998, 2004	Biology	Research article	Modals
Gledhill 2000	Medicine	Cancer research introductions	Phraseology
Marco 2000	Medicine	Research article	Frames
Okamura & Shaw 2000	Engineering	Transactional letters	Lexical phrases
Vassileva 2001	Linguistics	Research articles	Hedging and boosting
Hewings & Hewings 2002	Business	Journal articles and dissertations	Anticipatory <i>it</i> with extraposed subject
Hyland 2002a	Biology, physics, engineering, philosophy, sociology, marketing, applied linguistics	Journal article	First person pronouns
Hyland 2002c	Biology, physics, engineering, philosophy, sociology, marketing, applied linguistics	Research article	Reporting verbs
Silver 2003	History and economics	Research article	<i>Evidently</i>
Cortes 2004	History, biology	Research article	Lexical bundles
Hyland & Tse 2005	Applied linguistics, biology, business, computer science, engineering, public administration	Abstracts	Evaluative <i>that</i>
Harwood 2005	Business & Management, computing science, economics, physics	research article	First person pronouns
Groom 2005	History, literary criticism	Research article, book review	Grammar patterns involving <i>it</i>
Fløttum, Dahl, & Kinn 2006	Economics, linguistics, medicine	Research article	Pronouns, negation, discourse verbs, conjunctions
Hyland 2008	Engineering, biology,	Research articles	Lexical bundles

	business, applied linguistics		
Diani 2008	Linguistics, history, economics	Research article, book review article	<i>Really</i>
Aktas & Cortes 2008	Art and design, biology, computer science, economics, engineering, physics and astronomy	Research article	Shell nouns
ElMalik & Nesi 2008	Medicine	Research article	Modal verbs, probability adverbs, derivative adjectives, lexical bundles
Sheldon 2009	Applied linguistics and language teaching	Research articles	First person forms
Vongpumivitch, Huang, & Change 2009	Applied linguistics	Research article	AWL
Pecorari 2009	Biology	Research article	Lexical bundles
Bondi 2009	History, biology	Journal article	Temporal units
Bloch 2010	Biology, engineering	Critical review, research proposal	Reporting verbs
Gray 2010	Education, sociology	Research article	demonstratives
Gray & Cortes 2011	Applied linguistics, engineering	Research articles	<i>This</i> and <i>these</i>
Peacock 2011	Biology, chemistry, physics, environmental science, language and linguistics, law, business, public administration	Research articles	Introductory <i>it</i> plus that- and to-clause complementation

As illustrated in Table 3.1, corpus-based investigations on scholarly writing have shed light on a variety of lexicogrammatical properties in a variety of specific disciplines (22 reported here). Lexicogrammatical features investigated involve a range of structures; as mentioned in section 3.3.3 above, the most frequent involves phraseology (12 studies). Other structures investigated include verbs (10 studies), pronouns (5), adverbs (5), demonstratives (3), adjectives (3), nouns (2), negation (1), conjunctions (1), and prepositions (1). Functions investigated include hedging (6 studies), cohesion (4), identity representation (4), evaluation

(4), attribution (3), stance (2), emphasis (2), boosting (1), nominalization (1), causality (1), and relation of information (1). The inclusion of identity here was surprising, as remarked in section 3.2.3. As would be expected, findings involving these lexicogrammatical features and functions typically concern frequency counts. A range of disciplines representing both the sciences and humanities is also investigated in the selected studies presented in Table 3.1; these include biology (12 studies), engineering (8), applied linguistics (7), medicine (7), history (6), physics (5), business (5), economics (5), linguistics (4), computer science (3), sociology (3), philosophy (2), marketing (2), public administration (2), education (1), astronomy (1), chemistry (1), literary criticism (1), art and design (1), environmental science (1), law (1), and ecology (1).

It is hard to say which of the aspects—lexicogrammatical feature, function, or discipline—may have served as primary motivation for the studies (and thus aligning them with corpus linguistics as a discipline or method as discussed in section 3.3.2 above). For studies investigating phraseology, it may be reasonable to suppose that the interest in phraseology itself was a primary motivator, and thus these studies would be classified as belonging to the discipline of corpus linguistics. Given the role of some of the studies in L2 writing research, it may also be logical to presume function may have been a primary motivator, as in identity, and thus these studies would be classified as using a corpus-based approach to investigate issues in another discipline. Perhaps one way to make such a distinction is to consider the title of the studies. For example, Peacock (2011) investigated “introductory *it*” as a tool of evaluation in research articles in biology, chemistry, physics, environmental science, language and linguistics, law, business, and public administration. The title of the article “A comparative study of *introductory it* in research articles across eight disciplines” may mean that the investigation stemmed from a need to know or curiosity about

introductory *it* and how it varies according to discipline, rather than about evaluation (and thus classify it as corpus linguistics discipline). On the other hand, ElMalik and Nesi (2008) investigated modal verbs, probability adverbs, derivative adjectives, and lexical bundles for hedging and nominalization in research articles in medicine. The title of the article “Publishing research in a second language: The case of Sudanese contributors to international medical journals” may mean that the teaching of L2 writing in a specific context motivated the study (and thus classify it as corpus-based method). Though the idea of the motivating aspect is not particularly relevant to the research here, it is interesting to consider given the debates about corpus linguistics as a discipline in section 3.3.2, as well as the potentially competing genre- and corpus-based approaches (discussed in sections 3.4.2 above and 3.5.4 below). Considering the title of this thesis, “A Staircase Model for Teaching Grammar for EAP Writing in the IEP: Freshman Composition and the Noun Phrase” however, does not help to understand the primary motivation for the research reported here; this title addresses grammatical feature as well as the teaching of L2 writing in a specific context. Perhaps this means that the questions proposed here involving dichotomies—e.g. corpus linguistics as field or method—are not particularly important considerations.

One interesting feature of the studies in Table 3.1 is their tendency to compare findings, across discipline, register, level, and language. Half of the studies compare findings in at least two disciplines, some as many as eight (Biber, Conrad, & Reppen, 1998; Hyland, 2002a, 2002c, 2008; Silver, 2003; Cortes, 2004; Hyland & Tse, 2005; Harwood, 2005; Fløttum, Dahl, & Kinn, 2006; Diani, 2008; Aktas & Cortes, 2008; Sheldon, 2009; Bondi, 2009; Bloch, 2010; Gray, 2010; Gray & Cortes, 2011; Peacock, 2011). A portion of the articles compare findings across register (spoken versus written; e.g. Biber, Conrad, & Reppen, 1998; Cortes, 2004; Diani, 2008), as well as level of writing (graduate versus

scholarly, e.g. Hewings & Hewings, 2002; Hyland & Tse, 2005; Hyland, 2008; Akas & Cortes, 2008; Bloch, 2010); because findings in the latter do contribute to a description of both levels of writing, these publications have been included in both discussions here. For example, Hewings and Hewings (2002) investigate anticipatory *it* clauses with extraposed subject in business research articles as well as MBA dissertations; as such Hewings and Hewings (2002) are included in Table 3.1 as well as Table 3.2 below. Additionally, three studies compare their findings across languages—Vassileva (2001) examines hedges and boosters in linguistic research articles in English and Bulgarian; Fløttum, Dahl, and Kinn (2006) study a range of linguistic features across English, French, and Norwegian; Sheldon (2009) investigates first person forms in applied linguistics and language teaching research articles in English and Spanish. Finally, it is interesting to note that three studies employ a mixed methods approach—Hyland (1998, 2004) uses both corpus-based and ethnographic methods, and Pecorari (2009) uses both genre- and corpus-based methods.

The studies included here represent a selection of available research aimed at understanding academic writing in order to inform pedagogy to help students enter the discourse community.

3.5.3 Graduate writing (step 3)

While scholarly writing is expert writing, graduate writing is the penultimate step on the staircase model presented in Chapter 2. Graduate writers are often refining their discipline-specific knowledge in order to become an active part of their discipline's academic community. Although the number of studies investigating graduate level writing is considerably fewer than those focusing on scholarly writing, research on graduate writing, including masters' and doctoral dissertations and theses, also regularly involves understanding lexicogrammatical features in specific disciplines. An overview of the research

on graduate writing is provided in Table 3.2 and discussed below. As with scholarly writing, only work which specifies discipline, genre, and grammatical feature is included; for example, Kwan's (2006) investigation of the literature review in applied linguistics doctoral theses examines only the discourse moves of the section without reporting on any lexicogrammatical findings, and, therefore, it is not included here. The list in Table 3.2 is by no means exhaustive, but attempts to present a selection of available research.

Table 3.2 Selected Studies on Graduate Writing (Step 3)

<i>Author/Year</i>	<i>Discipline</i>	<i>Genre</i>	<i>Grammar feature</i>
Hewings & Hewings 2002	Business	dissertation	Anticipatory <i>it</i> with extraposed subject
Charles 2003	International politics, materials science	Thesis	Nouns
Hyland 2004	engineering, computer science, business, biology, applied linguistics, public administration	Thesis, dissertation	Conjunctions, modals
Hyland & Tse 2004	applied linguistics, biology, business, computer science, engineering, public administration	Acknowledgements in thesis and dissertation	Nominalization, performance verb, adjectives
Samraj 2004	Environmental science	Research paper	Sentence subject patterns
Hyland & Tse 2005	Applied linguistics, biology, business, computer science, engineering, public administration	Abstract	Evaluative <i>that</i>
Charles 2006	International politics, materials science	Thesis	Complement clauses
Koutsanti 2006	Engineering	Thesis	Personal pronouns
Starfield & Ravelli 2006	History, sociology	Thesis	First person pronoun
Charles 2007	International politics, materials science	Thesis	Nouns, complement clauses
Loudermilk 2007	Business	Thought essay	Pronouns, verbs
Aktas & Cortes 2008	Art and design, biology, computer science, economics, engineering, physics and astronomy	Research paper	Shell nouns

Hyland 2008	Engineering, biology, business, applied linguistics	Dissertation, thesis	Lexical bundles
Samraj 2008	Biology, philosophy, linguistics	Introduction in dissertation	First person pronouns
Isik Tas 2008	English language teaching	thesis	First person pronoun
Charles 2009	International politics, materials science	Thesis	Restrictive Adverbs
Flowerdew & Forest 2009	Applied linguistics	Literature review in thesis	keywords
John 2009	Applied linguistics	Dissertation	First person pronoun
Charles 2011	International politics, materials science	Thesis	Adverbials of result
Durrant & Mathews-Aydmh 2011	Economics, business, anthropology, law, tourism management, politics, publishing, sociology	Essay	<i>Wh-</i> structures

Similar to the research on scholarly writing, corpus-based investigations of graduate writing have also examined a variety of lexicogrammatical features in a variety of specific disciplines (also 22 here), including business (7 studies), biology (6), applied linguistics (6), engineering (6), international politics (6), materials science (5), computer science (4), public administration (3), philosophy (2), sociology (2), economics (2), history (1), English language teaching (1), linguistics (1), environmental science (1), art and design (1), physics (1), anthropology (1), publishing (1), law (1), tourism management (1), and astronomy (1). Lexicogrammatical features examined include pronouns (6 studies), nouns (4), phraseology (3), verbs (3), adverbs (2), complement clauses (2), conjunctions (1), *wh-* structures (1), and demonstratives (1). Investigated functions include identity construction (5 studies), stance (3), hedging (2), attribution (2), negotiation (2), gratitude (1), result/inference (1), cohesion (1), evaluation (1), indication (1), and emphasis (1). While research on scholarly writing is characterized by comparing findings across discipline, level, language, and method, studies

investigating graduate level writing, while comparing findings across several disciplines, tend to have a more singular focus overall.

3.5.4 Upper-level undergraduate writing (step 2b)

Step 2 of the staircase model presented in Chapter 2 is undergraduate writing. Undergraduate writing is divided into two substeps: first year undergraduate writing, including freshman composition (step 2a) and, as undergraduates become introduced to their chosen discipline and undertake discipline specific courses, upper-level undergraduate writing (step 2b). Step 2b is discussed here, while step 2a is discussed in section 3.5.5.

As mentioned in section 3.3.3, the development of register-specific freely available corpora, and publications stemming from them, are leading to fuller descriptions of academic registers. With the newly available MICUSP⁵ and BAWE⁶ Corpus, both of which are comprised of upper-level undergraduate writing⁷, we can expect rapid growth in the amount of research focusing on this level of writing using these resources; in fact, all the studies in this section make use of the BAWE or MICUSP with the exception of two, and they have been conducted only in the past three years. Table 3.3 provides an overview of the research in upper-level undergraduate writing. As above, only work specifying discipline, genre, and grammatical feature is included here; for example, Bruce's (2010) study of essays in English and sociology (taken from BAWE) examines only the rhetorical purposes and discoursal and text features of the essays without reporting any lexicogrammatical findings, and, therefore, it is not included in this discussion. This is an important point for this level of writing; the

⁵ *Michigan Corpus of Upper-level Student Papers*. (2009). Ann Arbor, MI: The Regents of the University of Michigan.

⁶ BAWE was developed at the Universities of Warwick, Reading and Oxford Brookes under the directorship of Hilary Nesi and Sheena Gardner (formerly of the Centre for Applied Linguistics [previously called CELTE], Warwick), Paul Thompson (Department of Applied Linguistics, Reading) and Paul Wickens (Westminster Institute of Education, Oxford Brookes), with funding from the ESRC (RES-000-23-0800). More details can be found at the corpus website: <http://www2.warwick.ac.uk/fac/soc/al/research/collect/bawe/>.

⁷ All years of undergraduate study are included in the BAWE corpus; however, because the British university system has no general education or freshman composition equivalent, writing in the BAWE has been classified as step 2b because it generally takes place in more discipline-specific courses.

number of investigations into upper-level undergraduate writing is already small, but those utilizing a corpus-based approach are fewer still⁸. An attempt has been made to include as many studies of level 2b writing as possible; the collection of work presented here is fairly exhaustive, although some studies may have been inadvertently excluded.

Table 3.3 Studies on Upper-Level Undergraduate Writing (Step 2b)

<i>Author/Year</i>	<i>Discipline</i>	<i>Genre</i>	<i>Grammar feature</i>
Cortes 2006	History	Response paper	Lexical bundles
Gardner 2008	History, engineering	Essay, report, funding proposal, reflective journal	Multi-dimensional analysis
Baratta 2009	Education	Essay	Passive
Gardezi & Nesi 2009	Economics, sociology, politics	Essays	Conjunctive adjuncts
Gardner 2009	Arts and humanities, life sciences, physical sciences, social sciences	Critique	Multi-dimensional analysis
Holmes & Nesi 2009	History, physics	Case study, critique, design specification, empathy writing, literature survey, methodology, narrative, essay, proposal, research report, proposing a solution	Keyword
Hyland 2009	Biology, engineering, information systems, business, economics, TESOL, public administration, social science	Final year project report	Questions, pronouns, directives,
Nesi 2009	Arts and humanities, life sciences, physical sciences, social sciences	Case study, critique, design specification, empathy writing, literature survey, methodology, narrative, essay, proposal, research report, proposing a	Multi-dimensional analysis

⁸ It might be interesting to note that those studies investigating 2b writing but not included in Table 3.3, many of them making use of BAWE, applied a genre approach (Bruce, 2010; Francis, Robson, & Read, 2010; Kusel, 1992; Robson, et al, 2010; Moore & Morton, 2005; North, 2005). See section 3.4.2 for a discussion of genre.

		solution	
Römer 2009	Linguistics, Philosophy, Psychology, and Sociology	Argumentative essay, creative writing, critique, proposal, report, research paper, response paper.	introductory <i>it</i>
Wulff & Römer 2009	Linguistics, Philosophy, Psychology, and Sociology	Argumentative essay, creative writing, critique, proposal, report, research paper, response paper.	Progressive aspect
Chen & Baker 2010	Arts and humanities, life sciences, physical sciences, social sciences	Case study, critique, design specification, empathy writing, literature survey, methodology, narrative, essay, proposal, research report, proposing a solution	Lexical bundles
Römer & Wulff 2010	Biology, Engineering, Classical Studies, Economics, Education, English, linguistics, Environment, Nursing, Philosophy, Physics, Political Science, Psychology, Sociology	Argumentative essay, creative writing, critique, proposal, report, research paper, response paper.	Attended and unattended <i>this</i>

As with scholarly and graduate writing, the studies investigating level 2b writing specifically consider discipline. The investigations at this level, though, are more similar to scholarly writing in that they compare findings across discipline. Unlike scholarly and graduate writing, however, there is more variation in the genres studied; this typifies the greater range of [some (e.g. Johns, 1997) would argue unique or pedagogical] genres represented in the university setting. As with discipline, findings are often compared across

genres. The studies in Table 3.3 using the MICUSP (Römer, 2009; Wulff & Römer, 2009; Römer & Wulff 2010) pinpoint individual linguistic features much like those in scholarly writing; some studies using BAWE (Gardner, 2008; Gardner, 2009; Nesi, 2009; Chen & Baker, 2010), however, have also concentrated on providing an overall description of the writings in the corpus by applying Biber's (1988) multi-dimensional analysis. Overall, research involving step 2b writing is, like that above, aimed at understanding academic writing, with an emphasis on informing pedagogy.

3.5.5 First-year undergraduate writing (step 2a)

First year undergraduate writing is step 2a, where graduates of North American high schools often begin the development of their academic writing skills in higher education, typically through freshman composition. As shown in Table 3.4, there are relatively few studies investigating this level of writing. Although only work which specifies discipline, genre, and grammatical feature is included in the discussions above, nearly all of the research focusing on writing in step 2a is included here. Whereas the final column in Tables 3.1-3.3 identifies the grammatical element examined in each study (thereby distinguishing these studies from others in that level of writing), the final column in Table 3.4 simply identifies research question addressed or approach used in the study, in order to account for all studies in this level of writing. An attempt has been made to include as many studies of first year undergraduate writing as possible; the collection of work presented here is fairly exhaustive, although some studies may have been inadvertently excluded.

Table 3.4 Selected Studies on First Year Undergraduate Writing (Step 2a)

<i>Author/Year</i>	<i>Discipline</i>	<i>Genre</i>	<i>Research Inquiry/Approach</i>
Ferris 1994	Freshman composition	Persuasive essay	Use of rhetorical strategies
Cortes 2002	Freshman composition	Description, research proposal, rhetorical analysis, research	Lexical bundles

		paper	
Hewings 2004	Geography	Essay	Rhetorical functions
Ravelli 2004	Management, history	Argumentative	Role of hyper-themes
Starfield 2004	Sociology	Essay	Identity
Soles 2005	Freshman composition	essay	Prose style chart
Keck 2006	English composition	Summary	Paraphrasing as a textual borrowing strategy
Xie, Ke, & Sharma 2008	Psychology	blog	Reflective thinking development
Shaw 2009	Literary studies	essays	Linking adverbials
Lee, S. 2010	First year writing	Persuasive essay	SFL—commands

Consistent with upper-level undergraduate writing, research focusing on first year undergraduate writing covers a wider variety of genres than that at the graduate or scholarly level; this illustrates a trend in writing level: the more expert the writing (the higher to the top of the staircase), the less varied the genre; conversely, the more novice the writing (the lower on the staircase), the more varied the genre⁹. Another difference between studies focusing on higher steps in the staircase model and that of 2a is the range of disciplines covered, and comparisons between them; while upper-level undergraduate, graduate, and scholarly writing cover as many as 40 disciplines, first year undergraduate writing focuses on only seven specific disciplines. This may be due to the small number of studies conducted, but may also be because the nature of first year undergraduate writing is such that it can only be discipline specific as far as freshman composition itself is a discipline (as claimed in Chapter 2). These trends support the staircase model proposed in chapter 2¹⁰; teaching a variety of genres¹¹ in one discipline is authentic pedagogy for EAP students in intensive English programs (IEPs).

⁹ Chapter 4 discusses the large number of text types comprising the student writing genre.

¹⁰ These trends also contribute to eliminating the “what should we teach” debate, at least at the EAP level.

¹¹ In Chapter 4 the term “text type” is used to classify various types of writing within the one “genre” of student writing; thus, a variety of text types is illustrated there.

Just from the brief review of literature at each of the steps on the staircase model—from expert to novice—provided in this section, there is evidence of a difference between scholarly writing and first year undergraduate writing, confirming the need for a fuller description of first year undergraduate writing. However, of the (merely) 10 studies investigating writing at step 2a, only 20% use a corpus-based approach; that is, only 2 of the studies have aimed to understand lexicogrammatical features which characterize first year undergraduate writing! Because freshman composition should be the model for teaching EAP in the IEP, as discussed in Chapter 2, the lack of research focusing on first year writing is both striking and lamentable. This is discussed more in section 3.5.7 below.

3.5.6 Learner writing (step 1)

Step 1 on the staircase model introduced in Chapter 2 is learner writing; students who are studying in EAP programs to obtain the language skills necessary to begin step 2, undergraduate writing. As discussed in Chapter 2, step 1 is where EAP writing in the IEP takes place. A surprising number of corpus-based studies investigating learner writing can be found in the literature. This step may be the most researched level of writing, even more than that of scholarly writing, making the gap in the step immediately above still more striking. This mainly stems from the interest in understanding the differences between L1 and L2 writing to inform pedagogical approaches and materials. Although an understanding of learner writing is a valuable domain for corpus-based L2 writing research, because the research involved in this thesis focuses on undergraduate writing (and graduate or scholarly writing as a poor model), the literature in step 1 is not covered here in the same intensity as above. Instead, Tables 3.5-3.7 provide an overview of L2 writers and texts as presented in Hinkel (2011) based on an extremely thorough review of hundreds of studies in L2 writing.

Table 3.5 L2 Writers (Source: Hinkel, 2011, p. 527-528)

organize and structure discourse moves differently
--

utilize discourse moves and their contents differently and inconsistently, primarily due to the negative transfer of discourse structuring conventions across various cultures
construct or place thesis statements differently, as well as omitting them altogether
take a logically and conceptually different approach to rhetorical development, argumentation, persuasion, and exposition/narration
often neglect to account for counterarguments and to anticipate audience reactions
support their arguments and claims by means of statements of personal opinions and beliefs in lieu of more substantive information
significantly more often leave their argumentation unsupported
sequence ideas and explanatory information differently: the norms of rhetorical structuring of discourse often do not conform to those expected in comparable written genres in English
construct less fluent and less detailed/explanatory prose
produce shorter and less elaborated texts
rely more on personal opinions and include less fact-based evidence in argumentation and exposition
over- or under-estimate the amount of readers' background knowledge and the need for textual clarity, explicitness, and specificity
differently orient the reader to the content, as well as differently introduce and develop topics; delay or omit thesis/main point statements and omit or dramatically shorten conclusions/closings
employ different strategies for extracting/citing information from sources, as well as paraphrasing, quoting, and including source material in their writing;
develop text cohesion differently, with weak lexical/semantic ties and theme connections, and a preponderance of overt discourse-level conjunctions;
rely on different given–new (theme–rheme) idea development;
use different sequencing, parsing, ordering, and connecting paragraph divisions, e.g., in some cases, such as those found in academic essays, L2 paragraphs need to be re-organized or divided into shorter ones, or short paragraphs need to be combined into longer ones;
differently—and often inconsistently—establish text cohesion: less frequent and less dense usage of cohesion devices, such as lexical, discoursal, and referential cohesive ties;
rely on repetition in order to paraphrase or establish cohesion at rates twice as high as those found in L1 writing;
develop prose that is oblique (e.g., hints) and vague (e.g., questions and allusions in lieu of direct statements);
often take moralistic and emotionally appealing approaches to argumentation and persuasion

Table 3.6 L2 Texts (Source: Hinkel, 2011, p. 529)

exhibit less lexical variety and sophistication
contain significantly fewer idiomatic and collocational expressions
have smaller lexical density and lexical specificity, and more frequent vocabulary misuses
rely on shorter sentences and clauses (aka T-units) with fewer words per clause and fewer words (e.g., nouns and modifiers) per verb
involve high rates of incomplete or inaccurate sentences (e.g., missing sentence subjects or verbs, incomplete verb phrases, sentence fragments)
repeat content words more often (i.e., nouns, verbs, adjectives, and adverbs)
provide twice as many simple paraphrases or avoid paraphrasing altogether with a

preponderance of referential pronouns
use shorter words (fewer words with two or more syllables)
use more conversational and high frequency words
incorporate fewer modifying and descriptive prepositional phrases, as well as a higher rate of misused prepositions
employ less subordination and two to three times more coordination
employ fewer passive constructions
employ fewer lexical (e.g., adjectives and adverbs) and syntactic modifiers (e.g., subordinate clauses) of sentences, nouns, and verbs
employ inconsistent uses of verb tenses
employ more emotive and private verbs
employ significantly higher rates of personal pronouns and lower rates of impersonal/referential pronouns
employ markedly fewer abstract and interpretive nouns and nominalizations
employ fewer adverbial modifiers and adverbial clauses
employ fewer epistemic and possibility hedges and more conversational hedges
employ more conversational intensifiers, emphatics, exaggeratives, and overstatements
employ fewer downtoners
employ more lexical softening devices

Table 3.7 Frequent Errors in L2 writing (Source: Hinkel, 2011, p. 530-531)

Word-level morphology (i.e., absent or incorrect affixes) and incorrect word forms	
Incomplete or incorrect subordinate clause structure (e.g., missing subjects, verbs and clause subordinators)	
Misuses (or under-uses and over-uses) of coherence and cohesion markers, such as coordinating conjunctions and demonstrative pronouns	
Singular or plural nouns and pronouns	Incorrect or omitted prepositions
Incorrect or omitted articles	Incorrect modal verbs
Verb tenses and aspects, and verb phrases	Subject and verb agreement
Sentence divisions, fragmented and clipped sentences, and run-ons	Spelling errors

Pedagogical implications of the findings in Tables 3.5-3.7 are widespread. The most important findings to consider for pedagogical implications of the research reported here, and discussed in chapters 6 and 7, include

- employment of fewer lexical and syntactic modifiers of nouns;
- incorporation of fewer modifying prepositional phrases.

3.5.7 *The gap in the research*

As discussed in section 3.4, corpus-based L2 writing research is increasing; especially with new student corpora such as MICUSP and BAWE, the gap between the number of expert and student writing studies is dwindling. Nevertheless, little research, corpus or otherwise, has been conducted which details linguistic features and functions present in first year undergraduate writing, which university-bound EAP students must confront. As discussed in Chapter 2, descriptions of academic prose—scholarly writing—or even that of graduate writing may not be valid for undergraduate writing or freshman composition (Hamp-Lyons & Kroll, 1996; Ginther & Grant, 1996), and using scholarly or graduate writing as models may even hinder students' writing development (Bazerman, 2006). With only two corpus-based studies of step 2a writing, there is an obvious gap in the research.

Two probable reasons for this gap in research were discussed in Chapter 2, section 2.4.1: the notion that undergraduate writing is not necessarily writing for an authentic audience and achieves no true communicative purpose (O'Brien, 1995; Johns, 1997) and the difficulty of determining which texts gatekeepers may constitute as prototypical examples. Another underlying cause potentially contributing to the void in literature on undergraduate writing is the question of discipline-specific discourse versus general academic discourse, also discussed in Chapter 2. Scholarly writing is written by experts in a subject for an audience with technical background on the subject, and as such, presumably shares a common discipline-specific rhetoric. In support of this argument, Elbow (1998, p. 148), for example, maintains that there is no such thing as general academic discourse:

I can't tell my students whether academic discourse in English means using a lot of structural signposts or leaving them out, bringing in their personal reactions or leaving them out, giving evidence from the poet's

life for interpretation or leaving that out, referring to the class, gender, and school of others or leaving that out. Even if I restrict myself to composition studies, I can't tell them whether academic discourse means quantitative or qualitative research or philosophical reflection. In short it's crazy to talk about academic discourse as one thing. (Elbow, 1998, p. 151)

Elbow (1998, p. 154) further argues that because the discourse instructors expect from students differs not only from instructor to instructor but also assignment to assignment, it is impossible to tell students what stylistic features are characteristic of the writing in a given discipline; Kusel's (1992) study of the structuring of essay introductions and conclusions across six subject departments supports this claim, as his results suggest that the rhetorical organization of these sections of essays is influenced significantly by the conventions adopted by the subject departments.

Dudley-Evans (1993), however, argues against the "common-core" teaching approach to EAP citing significant differences in two science disciplines, which more recent studies have confirmed. Spack (1998, p. 86) also questions the "common-core" teaching approach to EAP, arguing that L2 writing researchers and teachers should leave the teaching of writing in the disciplines to the instructors of those disciplines, as the goal of an EAP writing program should be to prepare students to become better academic writers in general; she does acknowledge, however, that determining what academic writing is and what ESL students need to know in order to produce it is no easy task. As Johns (1988, pp. 55-56) argues, though some generalizations can be made about the conventions and skills in academia, the differences among them may be greater than the similarities; for discipline, audience, and context significantly influence the language required.... We

are still having difficulty identifying the skills which are actually transferable to a variety of academic contexts.

It is no surprise, then, that Coxhead and Byrd (2007, p. 130) note that many L2 writing instructors are unsure how to provide students with the linguistic resources needed for successful academic writing, and Hinkel (2011, p. 535), almost a quarter century after Johns, remarks “it is not known what L2 writers are to be taught to enable them to meet their academic, occupational, professional, and vocational goals.” However, Coxhead and Byrd (2007, p. 134) argue, as is demonstrated in Chapter 5 here, that general academic writing comprises various grammatical features collaborating to create discourse; specifically, “...academic prose requires a cluster of grammatical items all working together; students need to learn to handle the whole set of characteristic vocabulary and grammar within the context of creating appropriately worded academic prose” (Coxhead & Byrd, 2007, p. 134). Similarly, it is this argument with which this study concurs, offering learners one piece to the puzzle: the noun phrase in freshman composition.

3.6 Summary

Because the next step for L2 writers seeking to engage in undergraduate studies in North America is step 2a, first year undergraduate writing, this thesis argues that materials and methods for English for academic purposes (EAP) writing instruction in the intensive English program (IEP) should be heavily influenced by freshman composition as it is taught in North American higher education contexts. This chapter presented an overview of research in the fields of L2 writing and corpus linguistics, including a discussion of how those fields have begun to interact. Debate surrounds the discussion of theory and method in both fields; L2 writing lacks a single, comprehensive theory, and researchers in corpus linguistics can not agree on its status as theory, discipline, or method. Yet, research in both fields has also been

expanding, and as L2 writing research has shifted to a focus on linguistic features of student texts, a natural convergence with corpus-based methods has been propelled by the desire to investigate features of academic writing and bring those discoveries to the classroom; this, in turn, has led to studies focusing on descriptions of the written academic register. Specifically, research conducted in the contexts appearing on the staircase model of progression in academic writing skills introduced in Chapter 2—scholarly writing, graduate writing, upper-level undergraduate writing, first year undergraduate writing, and learner writing—was reviewed in order to establish a gap in the research. With only two studies contributing to a lexicogrammatical description of first year writing, there is an obvious gap in the literature. Chapter 4 provides an overview of the corpora used and Chapter 5 discusses the methods used in this thesis to address that gap.

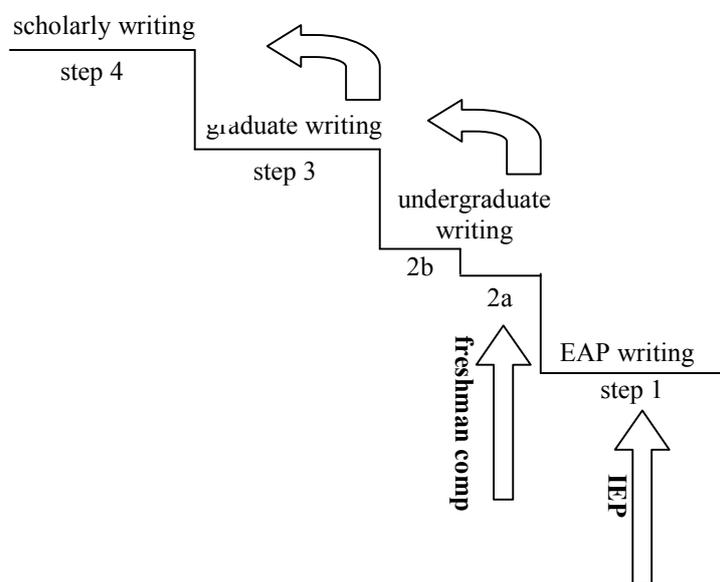
CHAPTER 4

THE CORPORA: FRESHMAN COMPOSITION WRITING AND SCHOLARLY WRITING

4.1 Introduction

Though much corpus-based L2 writing research in the past two decades has contributed to a description of the academic written register, the majority of this research has investigated expert writing, which is a poor model for novice writers (Hyland & Milton, 1997; Lorenz, 1999; Bazerman, 2006; Hüttner, 2008). To address the deficiency of the “expert writing” model, Chapter 2 proposed a staircase model to inform English for academic purposes (EAP) writing instruction in the intensive English program (IEP) (Figure 2.3 reproduced from Chapter 2 below). This model advocates teaching writing to those learners on step 1 of the staircase using a model informed by writing composed on step 2 of the staircase, specifically step 2a. As discussed in Chapter 3, there is a gap in the research investigating first year writing; therefore, this study compares the use of the noun phrase in freshman composition writing and scholarly writing with two main aims: to provide empirical evidence of the differences between the two levels of writing and to contribute to a description of first year undergraduate writing.

Figure 2.3 A staircase model of the progression of academic writing development in higher education (as shown on p. 40)



This chapter provides a description of the corpora used to compare the use of the noun phrase in scholarly writing and freshman composition writing. Three general academic writing corpora informed the description of scholarly writing used in this study. Together, the three corpora contain approximately 34 million words of scholarly writing from a variety of academic disciplines; these corpora are discussed in section 4.2. The North American Freshman Writing Corpus (NAFWiC), created for and used in this study to inform the description of first year undergraduate writing, contains nearly 250 texts specifically written in freshman composition courses; the NAFWiC is introduced and described in section 4.3.

4.2 Corpus-Cited References

4.2.1 Introduction

Corpus-cited references are extensive grammar and vocabulary books that cite corpus findings.¹ The two large corpus-cited references discussed below were discussed in Chapter 3 regarding the contributions they have made to corpus-based research. Using these two large

¹ A complete description of corpus-cited references, including methods for understanding and using them, is given in Chapter 5.

corpus-cited references available to date—*The Longman Grammar of Spoken and Written English (LGSWE)* (Biber et al., 1999), and the *Cambridge Grammar of English (CGE)* (Carter & McCarthy, 2006)—along with a smaller corpus-based reference—*A Communicative Grammar of English (ComGE)* (Leech & Svartvik, 2002)—a lexicogrammatical profile of scholarly writing (presented in Chapter 5) was compiled from a survey of the linguistic features which were identified as markers of scholarly writing. Selected features of the lexicogrammatical profile were then investigated in the NAFWiC (also discussed in Chapter 5).

4.2.2 *The Longman Grammar of Spoken and Written English*

Using a nearly 20 million word corpus with four main registers—conversation, fiction, newspaper language, and academic prose—as its source, the *LGSWE* provides a descriptive and empirical assessment of language in use that demonstrates the interface of grammar, vocabulary, and choice; choices involve various facets such as reason for communicating, context, audience, and mode (Biber, et al, 1999). Table 4.1 displays the number of tokens in the corpora used to inform the *LGSWE*. Using data from the corpora in Table 4.1, the *LGSWE* provides details of how, taken together, these choices give rise to systematic patterns of English (Biber et al., 1999 p. 4).

Table 4.1 Copora Used in the *LGSWE* (Biber et al, 1999, p. 25)

Core registers	Tokens
Conversation (BrE)	3,929,500
Fiction (AmE & BrE)	4,980,000
News (BrE)	5,432,800
Academic prose (AmE & BrE)	5,331,800
Total	19,674,100

Specifically, the academic prose subcorpus used to inform the *LGSWE* contains over 5.3 million words of general academic writing from American and British English in 408 texts (75 book extracts totalling 2,655,000 words and 333 research articles totalling 2,676,800

words) taken from a variety of disciplines in the sciences and humanities. Twelve of the disciplines present in the academic prose subcorpus are also investigated in the scholarly writing research discussed in Chapter 3 section 3.5.2; the academic subcorpus also includes an additional eight disciplines: agriculture, anthropology, archeology, geology, geography, mathematics, nursing, and psychology². Table 4.2 displays the number of tokens in the academic book extracts in the academic subcorpus according to discipline, and Table 4.3 provides the same for academic research articles. (In comparison, the NAFWiC, discussed in section 4.3, is larger than all except one of the 13 book extract subcorpora and all but three of the research articles subcorpora, putting it on par with the size of each of the subcorpora represented; in other words, in terms of size, the NAFWiC is sufficient to represent the freshman composition discipline if added to the academic prose subcorpus.) One key aspect to the scholarly writing corpus, however, is that nearly all the texts in the academic subcorpus were written for an audience with technical background on the subject, which does not include student textbooks (Biber et al., 1999, p. 32). This provides the main base of difference between the two corpora which merits their comparison.

Table 4.2 Academic Book Extracts according to Discipline (Biber et al, 1999, p. 33)

Subject	Tokens
Agriculture	179,000
Biology/ecology	190,200
Chemistry	158,200
Computing	269,300
Education	225,700
Engineering/technology	185,700
Geology/geography	152,200
Law/history	184,700
Linguistics/literature	149,600
Mathematics	216,600
Medicine	201,200
Psychology	118,400

² There are an additional nine disciplines investigated in the scholarly writing research discussed in Chapter 3 that are not included in the *LGSWE* academic subcorpus: applied linguistics, business, economics, philosophy, marketing, public administration, astronomy, art and design, and environmental science.

Sociology	424,200
Total	2,655,000

Table 4.3 Academic Research Articles according to Discipline (Biber et al, 1999, p. 34)

Subject	Tokens
Agriculture	179,000
Anthropology/archeology	152,100
Biology/entomology	369,100
Chemistry/physics	31,700
Computing	29,700
Ecology	13,100
Education	410,600
Geology	39,400
Law/history/politics	189,200
Linguistics	58,800
Mathematics	33,100
Medicine	752,000
Nursing	75,200
Psychology	124,100
Sociology	320,000
Total	2,676,800

It should be noted that the large majority of the findings presented in Chapter 6 are derived from the *LGSWE*.

4.2.3 Cambridge Grammar of English

Using the 1-billion-word plus Cambridge International Corpus (CIC) as its source, the *CGE* presents probabilistic rules of grammar, meaning that the rules presented “state what is most likely or least likely to apply in particular circumstances” (Carter & McCarthy, 2006, p. 5). Table 4.4 displays the number of tokens in the CIC according to register. The *CGE* provides a descriptive approach to grammar based on observation of usage, and the authors stress that both grammar as structure and grammar as choice are treated in the *CGE*, and the grammar of choice is as important as the grammar of structure (Carter & McCarthy, 2006, p. 7). According to Carter and McCarthy (2006), while it is not necessary for learners to mimic native speakers’ use of language to be successful, it is important for learners to observe and

understand how and why speakers use the language they do; describing language in use is not a prescription for learner use, but a presentation of data so that teachers and learners can make their own informed choices (Carter & McCarthy, 2006, p. 10), a significant aspect of language learning, and one that is revisited in Chapter 7.

Table 4.4 Tokens in the CIC according to Register (Cambridge University Press, 2011)

Register	Tokens
Written (AmE & BrE)	975,000,000
Spoken (AmE & BrE)	48,000,000
Written Academic (AmE & BrE)	29,000,000
Written Business (AmE & BrE)	100,000,000
Spoken Business (BrE)	1,000,000
Total	1,153,000,000

Though the *CGE* does not specify the details of the corpus used in the “Grammar and academic English” chapter of the reference, it does say that the chapter “focuses on items and structures which are common in academic language and which characterize it” (Carter & McCarthy, 2006, p. 267). It can be inferred, therefore, that the academic subcorpus of the CIC, a collection of American and British English texts from academic books and journals from a wide range of disciplines and topics comprising approximately 30 million words (Cambridge University Press, 2011), was used to report the findings in the *CGE*.

4.2.4 *A Communicative Grammar of English*

A Communicative Grammar of English (CommGE), now in its third edition, “has established itself as both an authoritative and an innovative grammar” (Leech & Svartvik, 2002, p. xi). The *CommGE* examines varieties of English—formal and informal, spoken and written—from a communicative rather than a structural approach to grammar. Though the *CommGE* is not a condensed version of *A Comprehensive Grammar of the English Language* (Quirk, Greenbaum, Leech, & Svartvik, 1985), it is partly based on that larger grammar (Leech & Svartvik, 2002). Importantly, the *CommGE* makes use of the Longman Corpus

Network to provide examples of authentic uses of English which illustrate grammatical statements in the reference. The Longman Corpus Network is a database of 330 million words from a wide range of authentic sources such as books, newspapers and magazines (Pearson Longman, 2011). More specifically, the Longman Corpus Network is comprised of five corpora:

- The Longman Spoken American Corpus comprises 5 million words transcribed from recordings representing the everyday conversations of more than 1000 Americans of various age groups, levels of education, and ethnicity, and includes speakers from over 30 US States (Pearson Longman, 2011).
- The Longman Written American Corpus comprises 100 million words from newspapers, journals, magazines, best-selling novels, technical and scientific writing, and coffee-table books; its design is based on the general design principles of the Lancaster Corpus (see below) (Pearson Longman, 2011).
- The Longman Learners' Corpus comprises 10 million words of language written by students of English. Every nationality and every language level is represented in the corpus (Pearson Longman, 2011).
- The Lancaster Corpus comprises 30 million words of written language taken from literature, magazines, papers and more ephemeral materials such as leaflets and packaging. The Lancaster Corpus claims to be the only global corpus that is carefully constructed to be as representative of written language as possible and a true reflection of twentieth century English (Pearson Longman, 2011).
- The Spoken British National Corpus comprises 10 million words of spoken English transcribed from natural, spontaneous conversations as well as from the language of

lectures, business meetings, after dinner speeches and chat shows (Pearson Longman, 2011).

Though the *CommGE* does not specify the details of the Longman Corpus Network used throughout the grammar, because of its distinction between formal and informal writing and speech it can be inferred that the technical and scientific writing from the Longman Written American Corpus, as well as other portions of the written corpora discussed above, were used to report the findings in the *CommGE* which refer to formal writing. While this may not be “academic language” specifically, the argument that academic language is formal writing would be difficult to refute.³

4.2.5 Summary

As will be detailed in Chapter 5, data based on the general academic writing corpora presented here and published in these three corpus-cited references (*LGSWE*, *CGE*, *CommGE*) was accessed to compile a lexico-grammatical profile of scholarly writing and, discussed in Chapter 6, to create a schematic representation of the modified noun phrase, which was in turn used to compare findings from the NAFWiC in order to determine differences between the two levels of writing.

4.3 The NAFWiC

4.3.1 Justification

The increasing number of studies focusing on discipline-specific writing, as discussed in Chapter 3, illustrates the interest in and usefulness of smaller, specialized corpora. Charles (2003, 2006, 2009) examines a number of grammatical features in a 500,000 word corpus of politics/international relations and materials science writing. Cortes (2006) investigates the effectiveness of teaching discipline-specific lexical bundles in an approximately 35,000 word

³ As will be discussed in Chapter 6, none of the findings in this study are based solely on the *CommGE*; it served primarily to confirm the data extrapolated from the *LGSWE* and *CGE*; thus the inference about the type of language used (“formal” as opposed to specifically “academic”) should not be a concern.

corpus of history essays. Pecorari (2009) has also investigated lexical bundles in a half a million word corpus of biology texts. Similarly, the discipline-specific subcorpora of the academic prose corpus used in the *LGSWE* average just under 200,000 words each (see Tables 4.2 and 4.3). While large corpora are certainly valuable, and necessary, for understanding overall trends and making generalizations about language as a whole, smaller, specialized corpora are more appropriate for informing EAP pedagogy (Flowerdew, 1993; Ghadessy, Henry, & Roseberry, 2001; Flowerdew, 2002; Tribble, 2002); this is largely an effect of EAP being “understood in terms of local contexts and the needs of particular students” (Hyland, 2006, p. 1) as discussed in Chapter 2 section 2.3.1. Undergraduate writing is the next step for learners in an EAP program, specifically those in intensive English programs (IEPs) with the aim of entering a North American college or university (see Figure 2.3 above); as such freshman composition—first year undergraduate writing—should inform instructional models for teaching EAP writing in the IEP. The NAFWiC is a small, specialized exemplar corpus, representing the discipline-specific writing of freshman composition compiled specifically for this study to contribute to such a model.

This section introduces and describes the NAFWiC, including development, design and organization, text type and assignment descriptions, representation, file names and annotation, and research and further development.

4.3.2 Development

I began compilation of the NAFWiC as part of this PhD thesis with the intent of gathering as many samples of proficient, freshman composition writing as possible during the first half of my doctoral studies. The NAFWiC consists of files from freshman composition courses, or their equivalent, from five institutions in the United States which were collected from May 2008 to December 2009, though some of the earliest writing samples were

composed in 1999. The uniqueness of the NAFWiC stems from its representation of a written register of academic English that has heretofore been underrepresented in corpus-based studies: first year undergraduate writing. As discussed in Chapter 3, a considerable body of work relating to scholarly writing is accessible (see Table 3.1), and while the newest released corpora focus on graduate or upper-level undergraduate writing (e.g. MICUSP, BAWE, see Table 3.2), there is a general lack of representation of undergraduate academic writing, particularly at the beginning undergraduate level⁴. Because of this, the NAFWiC is a specialized resource for the applied linguistics and TESOL professional communities. The corpus is useful for two main purposes. Firstly, the NAFWiC offers the practical application of materials writing for and teaching of grammar and writing to university-bound EAP students. As the NAFWiC contains examples of various types of writing that EAP students will need to have a good command of to succeed in their first year of undergraduate studies—nearly 250 authentic writing samples that received a grade of an A or B (merit or distinction equivalent)—it can be searched and utilized in countless ways for the study and teaching of writing skills (as an exemplar corpus, discussed in Chapter 2). Secondly, the NAFWiC is useful for research purposes as it can serve as primary data for quantitative and qualitative studies of discourse, which will further our understanding of the stylistic, lexical, and grammatical characteristics of this level of writing, contributing to an overall description of the academic written register. Questions such as how student writing differs from published writing or how first year writing differs from upper-level undergraduate writing, among others, can be explored.

⁴ As noted in Chapter 3, writing from all years of undergraduate study are included in the BAWE corpus; however, because the British university system has no general education or freshman composition equivalent, writing in the BAWE has been classified as step 2b because it generally takes place in more discipline-specific courses.

4.3.3 *Design and organization*

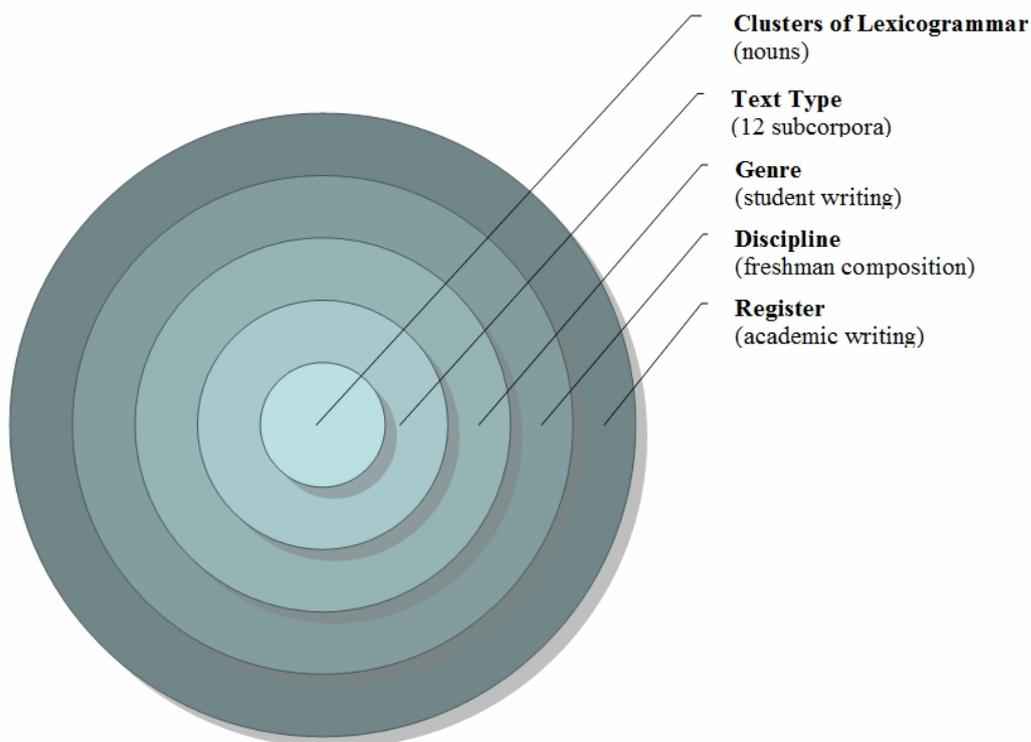
A corpus is a principled collection of authentic texts stored electronically (Hunston, 2002). As such, three principles guide the design of a corpus: (1) principled collection according to specific characteristics; (2) authentic texts used in a meaningful context; (3) electronic storage for access via computer software. Texts included in the NAFWiC are those that received an A or B (a mark of at least a 65% or a III equivalent) in an actual freshman composition classroom. Texts which merit this mark were selected because instructors perceived them as successful; this also allowed the decision of which texts constitute prototypical examples to be made by gatekeepers familiar with the disciplinary conventions, an important consideration for models discussed in Chapter 2. It should also be noted that each essay in the NAFWiC underwent multiple drafts as part of the composition process; it is the final draft, of course, that has been included in the NAFWiC. The NAFWiC is a stratified corpus, meaning that rather than trying to represent the proportions of student writing (based on enrolment numbers of students, for example), it tries to capture the essence of “good” writing, those assignments that received high marks as discussed above.

The NAFWiC is organized into 12 subcorpora, with each subcorpus representing a text type⁵. The use of the term “text type” here is nearly analogous to assignment type, though, as discussed below, different instructors may use different names for one type of assignment. A text type is defined according to the requirements of the assignment, and Bruce (2010) notes that text types can also be described in terms of the linguistic and stylistic features which characterize them. For example, the texts comprising the expository essay subcorpus were written in response to a particular assignment, which specifically required writers to present opinions on a topic, but also specifically instructed writers not to persuade

⁵ The organization model presented here is only one of several different options for organizing the NAFWiC. The corpus is still small enough that further study could organize the data differently in order to explore different avenues of research.

the reader for either side. The most well-known text-type is the essay. I assert here that lexicogrammatical structures cluster in text types⁶; text types comprise genres; genres are realized in disciplines; disciplines comprise the larger register of academic writing. These relationships are displayed in Figure 4.1.

Figure 4.1 Text Relationships in the NAFWiC



It should be noted that inside register, these relationships are not necessarily fixed, nor are the boundaries in the relationships rigid. Within the academic writing register, certain lexicogrammatical features are present in all text types, though there may be variation per text type. Some text types may be used in more than one genre; more than one discipline may use one particular genre. It is less likely, though, that a text type or genre from academic writing would be found in another register, and it is already clear that aspects of lexicogrammar cluster differently within different registers; thus, other registers have their own relationships.

⁶ The “cluster” of lexicogrammatical structures was first introduced in Chapter 3. It is explored further in Chapter 5.

This figure is similar to Stubbs' (2010) idea of how "words relate to the world", but takes the perspective how lexicogrammar and text types relate to the world. This is also one representation of "how writing is performed in a social act" (Hyland, 1994, p. 240). Inserting a social act into the second or third rings defines what may appear in the other rings.

Two arguments made in Chapter 2 are relevant to the discussion here:

1. Particular assignments are common and recurring in freshman composition, and over time they have become a genre, which uses lexicogrammatical features differently than other genres;
2. Freshman composition is defined as a discipline based on its courses, assignments, and distinctive realization of shared features in them.

Thus, here, 12 text types comprise the genre of student writing realized in the discipline of freshman composition, one contributor to the register of academic writing. (The lexicogrammatical features comprising the inner ring are discussed in Chapters 5 and 6.)

The corpus has been organized according to text type based on two considerations. The first consideration is the diverse nature of the freshman composition course. Understanding the various text types dealt with in freshman composition is one step to understanding the genre of student writing. Secondly, research already suggests lexical and grammatical variation among genres within the larger categories of discipline and academic writing (e.g. Hewings, 2006); one potential area of research to be explored through the NAFWiC is how much variation occurs within the text types themselves (e.g. personal narrative to film analysis) that make up the student writing genre. In addition, the NAFWiC can be searched according to text types which require the use of sources.

4.3.4 Text types and assignment descriptions

As mentioned above, the NAFWiC is organized into 12 subcorpora, with each subcorpus representing a text type. Table 4.5 displays a list of subcorpora, including number of texts, tokens, and types.

Table 4.5. NAFWiC Subcorpora

<i>Text type</i>	<i># of Texts</i>	<i>Tokens</i>	<i>Types</i>
Expository Essay	30	37,195	4,514
Persuasive Essay	40	55,611	6,493
Personal Narrative	24	28,390	3,055
Summary Synthesis	28	27,319	3,490
Rhetorical Analysis	32	39,972	3,898
Remembering an Event	7	7,616	1,840
Profile	10	9,788	2,253
Proposing a Solution Essay	31	77,208	8,189
Film Analysis	7	5,613	1,309
Literary Research Paper	5	8,205	1,865
Thematic Analysis	14	14,445	2,228
Character Analysis	18	17,277	2,804
TOTAL	246	328,639	41,938

Four instructors from five institutions contributed the 246 texts. Unsurprisingly, because of the diversity of the freshman composition course, not every instructor required the exact same assignment or even used the same name for a similar assignment; however, as discussed in Chapter 2 and noted above, given that particular assignments are common and recurring in freshman composition, similarities are apparent. The NAFWiC contains only 12 subcorpora, but 18 assignments appear in the NAFWiC. The main distinction between the “assignments” and the “text types” is the name given the assignment by the instructors. For example, Instructor B required an “expository” assignment while Instructor C required an “opposing views” assignment. Based on the assignment descriptions, however, both the expository assignment and the opposing views assignments are expository text types (i.e. they required writers to present opinions on a topic, but also specifically instructed writers not to persuade the reader for either side.). In addition to combining the expository and opposing

views assignments into the Expository Essay subcorpus, the persuasive, personal opinion, defending a thesis statement, critical response, argument, and extended research paper assignments were combined into the Persuasive Essay subcorpus; and the library research and proposing a solution assignments were combined into the Proposing a Solution subcorpus. The assignment description for each text type by subcorpus has been reproduced below to offer an understanding of the meaningful context of each text type.

4.3.4.1 Expository Essay Subcorpus

The expository essay subcorpus contains 30 files, just over 37,000 tokens and two assignments, the expository essay and the opposing views essay. Essays in this subcorpus were written in two different freshman composition courses. In Instructor B's class, the expository essay assignment asked that writers explain the facts about California Proposition 8 as clearly and accurately as possible to help the reader understand the background information in order to analyze the issue for him/herself. The assignment description specifically noted that the writer should not try to share his/her own opinion or to persuade the reader on the issue. For instructor B's class, the expository essay required a minimum of 1,200 words and six to eight sources. For instructor C's class, the opposing views essay assignment asked writers to provide information about a debatable topic without showing bias; elements of the essay were to include an introduction to the issues and each position, a comparison of the topics, a logical plan, and a fair and unbiased presentation of the issues. For instructor C's class, the assignment required between 1,000 and 1,200 words and the use of three to six sources.

In sum, the expository essay text type

- discusses a topic
- using multiple sources

- without bias.

4.3.4.2 Persuasive Essay Subcorpus

The persuasive essay subcorpus contains 40 files, over 55,000 tokens, and six assignments—persuasive essay, personal opinion essay, defending a thesis statement essay, critical response, argument essay, and extended research paper. The persuasive essay subcorpus is the largest subcorpus in terms of files and the second largest in terms of tokens. Essays in this subcorpus were written in three different freshman composition courses. In Instructor B’s class, the persuasive essay assignment asked that writers research and form opinions about the 2008 presidential election and write an essay explaining and supporting their opinions on the topic, clearly and persuasively explaining why they hold their opinions, as well as anticipating and countering possible points of disagreement; the assignment required a minimum of 1,500 words and the use of two to five sources.

In Instructor C’s class, the personal opinion assignment asked that writers form opinions about a topic of their choice and write an essay explaining and supporting their opinions on that topic. The assignment required approximately 1,500 words, and the use of sources was optional.

Instructor D assigned three progressively more difficult persuasive essays. In the first, and least difficult assignment, critical response, writers were asked to provide detailed and thoughtful reflections about a literary text; the assignment required specifically that writers state a point about some aspect of the text and support the point with evidence and analysis. This assignment required a minimum of 500 words and the use at least one source (presumably the literary text). In the second assignment, argument essay, Instructor D required writers to take a clear stand on any issue, presenting supporting arguments as well as points of refutation; this assignment required between 750 and 1,250 words, and the use of

sources was optional. The final, and most difficult, assignment, the extended research paper, required that writers advance a thesis statement and attempt to convince readers to agree with that thesis; this assignment required between 1,500 and 2,000 words and the use of six to eight sources.

In sum, the persuasive essay text type

- discusses a topic
- from two angles
- preferably using multiple sources
- lobbying one side.

4.3.4.3 Personal Narrative Subcorpus

The personal narrative subcorpus contains 24 files and more than 28,000 tokens. Texts in this subcorpus were written at two different institutions but under one instructor. The personal narrative required writers to use narration and description to show readers who they are and how they became that way; the personal narrative assignment specifically required the use of specific, concrete details to construct a mental image of the situation. The assignment required between 750 and 1,250 words and did not require sources.

In sum, the personal narrative text type

- uses a place, event, and person
- to define the writer
- without the use of sources.

4.3.4.4 Proposing a Solution Essay Subcorpus

The proposing a solution essay subcorpus contains 31 files, over 77,000 tokens, and two assignments, library research essay and proposing a solution essay. The proposing a solution essay subcorpus is the largest subcorpus in terms of tokens and the third largest in

terms of files. Texts in this subcorpus were written at three different institutions but under two instructors. In Instructor A's courses, the library research assignment asked that writers propose solutions for public issues through research and required between 2,500 and 3,000 words and the use of five to eight sources. In Instructor C's class, the proposing a solution assignment required writers to use personal knowledge and research to present a solution to a problem in today's society; elements of the assignment included a well defined problem, a clearly described solution, a convincing argument supporting the proposed solution, an anticipation of reader's objections, and an overview of alternative solutions. The assignment required between 1,500 and 2,000 words and the use of three to six sources.

In sum, the proposing a solution text type

- discusses a problem
- from two angles
- using multiple sources
- presenting a solution.

4.3.4.5 Summary Synthesis Subcorpus

The summary synthesis subcorpus contains 28 files and over 27,000 tokens. Texts in this subcorpus were written at two different institutions but under one instructor. The summary synthesis assignment required writers to present different points of view on the same topic by describing texts, their main points and supporting details, while organizing the ideas and examples in conjunction with their own view on the topic. The assignment required between 750 and 1,000 words and the use of a minimum of two sources.

In sum, the summary synthesis text type

- discusses a topic
- exemplified in supporting texts

- using more than one source.

4.3.4.6 Rhetorical Analysis Subcorpus

The rhetorical analysis subcorpus contains 32 files and nearly 40,000 tokens. Texts in this subcorpus were written at two different institutions but under one instructor. The rhetorical analysis assignment required writers to discuss an issue of public concern by presenting different aspects on the issue in a critical manner, specifically by breaking down literary texts on the topic to glean more specific understanding. The assignment required between 750 and 1,250 words and the use of at least one source.

In sum, the rhetorical analysis text type

- discusses a problem
- presented and exemplified in a literary text
- using only that text as a source.

The texts in the rhetorical analysis and summary synthesis subcorpora are similar in their use of outside texts to provide evidence for discussion points on a topic; they merit separate subcorpora, however, because of the specific requirement of only one source (a literary text) in the rhetorical analysis subcorpus and the requirement of at least two sources (of any type) in the summary synthesis subcorpus, as the combining and reporting of two or more sources requires the employment of different writing skills and linguistic features than only one source.

4.3.4.7 Remembering an Event Subcorpus

The remembering an event subcorpus contains 7 files and nearly 8,000 tokens. Texts in this subcorpus were written under one instructor. The remembering an event assignment required writers to write a narrative about a past event which included a well-told story, vivid

descriptions, and an indication of the event's significance. The assignment required between 1,000 and 1,200 words and no sources.

In sum, the remembering an event text type

- narrates an event
- discussing the importance of the event
- without the use of sources.

4.3.4.8 Profile Subcorpus

The profile subcorpus contains 10 files and nearly 10,000 tokens. Texts in this subcorpus were written under one instructor. The profile assignment required writers to observe a place for 45 minutes and profile it so a reader feels he/she has been there; elements of the essay included a description of people and place, background information, a topical or narrative plan, evidence of either a detached observer or a participant observer, and demonstration of perspective on the subject. The assignment required between 1,000 and 1,200 words and no sources.

In sum, the profile text type

- describes a place
- evoking feelings
- without the use of sources.

4.3.4.9 Film Analysis Subcorpus

The film analysis subcorpus contains 7 files and nearly 6,000 tokens. The film analysis subcorpus is the smallest subcorpus in terms of tokens and the second smallest in terms of files. Texts in this subcorpus were written under one instructor. The film analysis assignment required writers to critique the film *Wal Mart: The High Cost of Low Prices* by focusing on the arguments the film made, the evidence it used to support those arguments,

and the strengths and/or weaknesses of those arguments. The assignment required a minimum of 500 words and one source (presumably the film).

In sum, the film analysis text type

- presents a critique
- using evidence from the film
- using only the film as a source.

4.3.4.10 Literary Research Paper Subcorpus

The literary research paper subcorpus contains 5 files and more than 8,000 tokens.

The literary research paper subcorpus is the smallest subcorpus in terms of files, but only the third smallest in terms of tokens. Texts in this subcorpus were written under one instructor.

The literary research paper assignment required writers to reference the novel *One Flew Over the Cuckoo's Nest* to extend or support their own analysis; specifically, the assignment asked writers to tell the reader what they think about an issue and use the source to illustrate and support their opinion. The assignment required between 1,500 and 2,000 words and the use of five to eight sources.

In sum, the literary research text type

- discusses a problem
- presented and exemplified in a literary text
- using multiple sources.

The texts in the literary analysis subcorpus are similar to those in the rhetorical analysis and summary synthesis subcorpora but merit a separate subcorpus because of the specific requirement of the one literary source (rhetorical analysis) in combination with the requirement of multiple outside sources (summary synthesis).

4.3.4.11 Thematic Analysis Subcorpus

The thematic analysis subcorpus contains 14 files and more than 14,000 tokens. Texts in this subcorpus were written under one instructor. The thematic analysis assignment required writers to analyze a theme in either the short story “Trifles” or the novel *A Raisin in the Sun* and the elements the author uses to develop the theme. The assignment specified that although a work can have numerous themes, the writers were to focus on one dominant theme, preferably one that they could agree with or that with which they had experienced in their own life. The assignment required between 750 and 2,500 words and one source (presumably the referenced story).

In sum, the thematic analysis text type

- discusses a theme
- developed in a literary text
- using only that text as a source.

While essays in the thematic analysis and rhetorical analysis subcorpora are similar in their use of a literary text to discuss a topic, they merit separate subcorpora because of the different foci of the essays: a larger issue plaguing society in the rhetorical analysis subcorpus and the literary theme specific to the literary text in the thematic analysis subcorpus.

4.3.4.12 Character Analysis Subcorpus

The character analysis subcorpus contains 18 files and just over 17,000 tokens. Texts in this subcorpus were written under one instructor. The character analysis assignment required writers to analyze any literary character’s personality through the traits he or she exhibits, the changes he or she undergoes, or the methods the author uses to develop the character. The assignment required between 750 and 2,500 words and the use of one source (presumably the story in which the discussed character appears).

In sum, the character analysis text type

- discusses character traits
- developed in a literary text
- using only that text as a source.

4.3.4.13 Assignment Topics

As mentioned above, on occasion, a topic was assigned for a particular assignment. “California Proposition 8” was specifically assigned as the topic for the expository essay written in Instructor B’s class, and the “2008 Presidential Election” was specifically assigned as the topic for the persuasive essay written in Instructor B’s class. *Wal Mart: The High Cost of Low Prices* was the film specifically assigned for the film analysis essays written in Instructor D’s class, and *One Flew Over the Cuckoos Nest* was specifically assigned as the topic for the literary research paper written in Instructor D’s class. Additionally, for the thematic analysis essay written in Instructor D’s classes, students could choose topics from the two stories “Trifles” or *A Raisin in the Sun*. Table 4.6 provides an overview of assigned topics for applicable assignments.

Table 4.6 Assigned Topics

Assignment	Required Topic
Expository Essay	California Proposition 8
Persuasive Essay	2008 Presidential election
Film Analysis	<i>Wal Mart: The High Cost of Low Prices</i>
Literary Research Paper	<i>One Flew Over the Cuckoos Nest</i>
Thematic Analysis	“Trifles” or <i>A Raisin in the Sun</i>

4.3.5 Representation

The NAFWiC is representative of both the freshman composition course and first year undergraduate writing. Section 4.3.5.1 discusses the instructors and institutions in regards to the representativeness of the freshman composition course; section 4.3.5.2 discusses the size of the corpus as representative based on results of a pilot study.

4.3.5.1 Freshman Composition

4.3.5.1.1 Introduction

Using my own networking base, I reached out to instructors from the TESOL community via the second language writing interest section e-list, instructors from institutions where I currently serve as an adjunct professor, as well as personal friends who are involved in composition instruction in order to gain participation for creation of the NAFWiC. Four instructors teaching at five institutions in the United States submitted students' essays that received a mark of A or B in their composition course. See Table 4.7 for an overview of each instructor, institution, and course. During the course, instructors asked students who would be willing to contribute their work to the corpus to sign an informed consent form (see Appendix 2) which transferred copyright of the original work to the NAFWiC. Students were not compensated for their submissions, though two students requested letters thanking them for their "service to academia" for use in funding applications. Instructors submitted all files to me electronically.

Table 4.7 General Instructor, Institution and Course Information

<i>Instructor Profile</i>	<i>Institution Overview</i>	<i>Course Description</i>
Instructor A—PhD, 10 years experience	Large research institution in the southwest; large non-research institution in the south	focuses on critical reading and writing in the academic community; intensive instruction in writing processes
Instructor B—PhD, 20 years experience	Large research institution on the west coast	composition, the essay, paragraph structure, diction, and related topics
Instructor C—M.A., 5 years experience	Small private institution in the north	the study and practice of the process approach to writing with attention to the rhetorical dynamic of the writer, text, and reader
Instructor D—M.A., 10 years experience	Small community college in the south	principles and techniques of expository and persuasive composition, analysis of texts with introduction to research methods, and critical thinking

To further demonstrate the diversity involved in first year writing courses and the representation of that diversity in the NAFWiC, the following sections provide a comprehensive overview of the instructors, the courses, and the institutions from which the texts were submitted. This overview facilitates observation about five points of comparison for the instructors—education, field, experience, status, philosophy—and four points of comparison for the institutions—type, size, status, and region. The breadth of instructor experience and credentials as well as the diversity of settings exemplifies not only the diversity involved in the first year writing course, but also demonstrates how the essays in the NAFWiC come from courses generally representative of the context itself. These observations of diversity are displayed in Tables 4.8 and 4.9 at the end of this section.

Section 6.5 discusses a pilot study conducted prior to the commencement of the research reported in this thesis; as discussed in that section, the NAFWiC shows the same general trends as those of a different corpus of first-year undergraduate writing, offering further confirmation of the representativeness of the NAFWiC.

4.3.5.1.2 Instructor A

Instructor A holds a PhD in applied linguistics. She began teaching writing in 1999 as a postgraduate student, and her experience in teaching writing stems from college composition courses at a research university with a population of nearly 20,000 students in a southwestern state of the U.S., as well as a non-research university with a similar population size in a southern state of the U.S. The complete course description and goals for Instructor A's courses as cited in the course syllabi:

- English 105 Course Description

This course focuses on critical reading and writing in the academic community. Throughout the semester we practice the *reading process*: generating questions or deriving answers from texts; summarizing texts; identifying examples, drawing inferences, and making logical or comparative connections; organizing information in a variety of ways; seeing and learning rhetorical skills used by effective writers; and evaluating the merits of what we read. More urgently, we also practice the *writing process*: identifying audience and purpose; gathering or finding ideas; organizing and interrelating those ideas for readers; drafting in order to develop, support, and illustrate ideas; revising from trial-and-error and in light of peer input; editing for clarity and accuracy.

- English 105 Course Goals

- develop critical reading skills through close attention to text content and to the skills needed to interpret texts effectively
- develop expository writing skills through attention to the writing process
- apply critical reading and writing skills to formal writing tasks, including an extended writing project
- develop technological literacy skills to rhetorically analyze online resources based on the audience addressed, the purpose explored, and the language used

- English 1101 Course Description

Intensive instruction in writing processes. The course focuses on organization of ideas in well-developed expository and argumentative essays (usually six to eight essays), with stress on grammar, punctuation, and vocabulary development.

- English 1101 Course Goals

- understand and apply professional skills in manuscript editing, project management, peer feedback, and professional language use

- describe the content and organization of texts (e.g., memories, academic essays) and evaluate their relevance and usefulness to a specific task or project
- understand and apply writing processes (reading, inventing, drafting, revising, editing)
- understand and apply rhetorical ways of organizing writing (e.g., description, narration, comparison and contrast, cause and effect, classification, exemplification)
- understand and apply professional technology skills to access online library databases, participate in electronic discussions, and use other useful online resources.

Instructor A maintains an academic writing focus in the course, stemming from a genre-based approach as well as includes corpus-based activities in her class.

4.3.5.1.3 Instructor B

Instructor B holds a PhD in applied linguistics. She has been teaching writing for almost 20 years, and her experience in teaching writing stems from instruction of graduate MA TESOL courses and directing ESL and undergraduate writing programs at a research university with a population of nearly 25,000 students on the west coast of the U.S. The complete course description and course goals for Instructor B's course:

- University Writing Program (UWP) 1 Course Description

Composition, the essay, paragraph structure, diction, and related topics. Frequent writing assignments will be made.

- UWP 1 Course Goals

- develop the close reading skills necessary for analysis and interpretation of academic and scholarly writing
- introduce students to the concepts of audience, purpose, persona, voice, authority, and tone as they relate to expository writing
- introduce the forms and conventions of non-fiction prose

- explore, through readings, how assumptions, key questions, and fundamental concepts lead to the construction of knowledge in different disciplines
- explore the nature of evidence in academic and expository writing
- review the requirements of standard written English and to help students master accepted grammar, syntax, and usage
- develop students' ability to recognize the stylistic aspects of expository texts, and develop a clear, reasonably sophisticated, and appropriately varied prose style in their own writing
- develop students' awareness of language, including such concepts as diction, word choice, connotation/denotation, and figurative language
- introduce students to effective ways to structure and organize texts
- help students learn how to analyze individual arguments
- provide students with instruction and practice in synthesizing multiple texts, formulating an original argument, and supporting it with appropriate evidence

Instructor B also maintains an academic writing focus in the course; though she does not focus as much on the genre- or corpus-based approaches, she does give considerable attention to both product and process.

4.3.5.1.4 Instructor C

Instructor C holds an M.A. in education. She began teaching writing in 2002, and her experience in teaching writing stems from her positions as an adjunct instructor of English and the Writing Center Coordinator at a private university with a population of nearly 3,000 students in a northern state of the U.S. The complete course description and objectives for Instructor C's course:

- English 113 Course Description:

The study and practice of the process approach to writing with attention to the rhetorical dynamic of the writer, text, and reader. This course includes essays, the study of research, a library orientation, a review of reading strategies, and readings relevant to the study of written discourse. Grammar is studied as part of precise, effective written communication.

- English 113 Course Objectives:
 - use writing and reading for inquiry, learning, thinking, and communicating
 - develop strategies for generating, revising, editing, and proof-reading
 - write with an identifiable point or thesis
 - respond appropriately to different kinds of rhetorical situations
 - recognize and use a number of strategies for developing paragraphs
 - demonstrate understanding of effective writing choices and use of writing techniques such as parallelism and subordination
 - integrate own ideas with facts and opinions of others
 - conduct library research and appropriately use and document source material in writing

Instructor C maintains a general writing approach in the course with the belief that the skills the students learn in her course, namely those focusing on process, will transfer to other courses. The instructor does not have a background in applied linguistics, and although that is not a requirement for a focus on academic writing, it is more logical that an instructor with such a background would have a focus on academic writing, like Instructors A and B.

However, neither does this instructor have a background in the English literature field. Again, that is not a requirement for a focus on process and general writing skills, but it is logical that an instructor with such a background, such as Instructor D, would have a focus on those.

Although this general writing skills focus could be due to any number of reasons, perhaps it is the absence of specific language training which leads to it.

4.3.5.1.5 Instructor D

Instructor D holds a Master's in Liberal Arts degree. He began teaching writing in 1998, and his experience in teaching writing stems from college composition courses at a community college with a population of nearly 2,000 students in a southern state of the U.S.

The complete course description and objectives for Instructor D's courses:

- English 1013 Course Description

Principles and techniques of expository and persuasive composition, analysis of texts with introduction to research methods, and critical thinking.

- English 1013 Course Objectives

- construct a logical argument incorporating the rhetorical modes of composition
- write a unified, coherent, well-organized essay
- write a grammatically correct in-class essay
- engage successfully in all stages of the writing process
- write a research paper using MLA format

- English 1023 Course Description

Further studies of principles and techniques of expository and persuasive composition, analysis of texts with introduction to research methods, and critical thinking.

- English 1023 Course Objectives

- respond appropriately to various rhetorical situations, purposes, and audiences
- use writing and reading for inquiry, learning, thinking, and communicating
- integrate original ideas with those of others
- develop flexible strategies for generating, revising, editing, and proof-reading

- use collaborative writing processes
- demonstrate knowledge of structure, paragraphing, tone, mechanics, syntax, grammar, and documentation by constructing a coherent, organized analysis essay
- identify literary elements
- analyze elements of fiction, poetry and drama
- define discipline-specific terms
- distinguish characteristics of each of the literary genres
- research and write about a critical approach to literature

Instructor D also maintains a general writing approach in the course with the belief that the skills the students learn in his course, namely those focusing on process, will transfer to other courses. This view is not particularly surprising given the instructor's background in the English literature field, specifically creative writing.

4.3.5.1.6 Overview

Tables 4.8 and 4.9 provide an overview of diversity among instructors and institutions. Again, the breadth of instructor experience and credentials as well as the diversity of settings presented here demonstrate not only the diversity involved in the first year writing course, but also how writing in the NAFWiC represents that diversity. Additionally, these sections have demonstrated how the course descriptions, goals, and objectives resonate with the general description of freshman composition discussed in Chapter 2.

Table 4.8 Diversity among Instructors

Instructor	Education	Field	Experience	Status	Focus
<i>Instructor A</i>	PhD	Applied Linguistics	10 years	Post graduate student/full time instructor	Academic Writing
<i>Instructor B</i>	PhD	Applied Linguistics	20 years	Tenured Faculty	Academic Writing
<i>Instructor C</i>	M.A.	Education	5 years	Adjunct	General Writing Skills
<i>Instructor D</i>	M.L.A.	Creative Writing	10 years	Full time Instructor	General Writing Skills

Table 4.9 Diversity among Institutions

Institution	Type	Size	Status	Region
<i>Institution A1</i>	4 Year University	20,000	Public Research	Southwest
<i>Institution A2</i>	4 Year University	20,000	Public Non-Research	South
<i>Institution B</i>	4 Year University	25,000	Public Research	West
<i>Institution C</i>	4 Year University	3,000	Private Non-Research	North
<i>Institution D</i>	Community College	2,000	Public	South

4.3.5.2 Size

As discussed in section 4.3.1 above, small, specialized corpora have been used in studies focusing on discipline-specific writing; the size of the NAFWiC is comparable to such studies discussed in Chapter 3, as well as with the discipline-specific subcorpora discussed in section 4.2. Furthermore, the NAFWiC is representative of first year undergraduate writing based on the results of a pilot study conducted at the onset of the research for this thesis.

The pilot study employed many of the same methods discussed in Chapter 5, but used the Viking Corpus, a specialized corpus which aims to represent all levels of successful student writing from the major university disciplines, as data. Release 1 of the corpus,

portions of which were used for the pilot study, contains approximately 400 student writing samples of 11 types which received a B (mark of 2 equivalent) or higher in 26 disciplines for a total of just over 600,000 words. Those texts which were written by first year undergraduate students, indicated by a “100 level⁷” label in the VIKING corpus, were compiled into the 100s subcorpus creating a corpus exclusive of first year undergraduate writing. The 100s subcorpus, just over 73,000 words, comprises 70 texts of nine types written by 28 students, both native and nonnative speakers of English, from eight departments (Albers, 2007, p. 9). The major difference between the NAFWiC and the 100s subcorpus of the Viking Corpus is the eight different disciplines represented in the 100s subcorpus (as opposed to just freshman composition in the NAFWiC); the disciplines include Business Administration, Criminology, English, University Studies, History, Psychology, Political Science, and Philosophy.

A random sample (10%) of essays from the 100s subcorpus were analyzed for the same features of the noun phrase (discussed in Chapters 5 and 6) as those from the NAFWiC. Though the results from the 100s subcorpus and the NAFWiC are not identical, the same trends occur in both corpora. In fact, the “bullet point” findings reported in sections 1.3.4 and 6.4.3 regarding the NAFWiC are nearly identical to those from the 100s subcorpus analysis. General patterns of successful first year undergraduate writing across multiple disciplines (as represented in the the Viking corpus), different from scholarly writing, that emerged from the pilot study include

- overall greater use of modifiers,
- less reliance on the definite article,
- greater use of single adjective pre-modifiers, and

⁷ Indicating freshman level.

- more diverse use of post-modifiers.

These findings demonstrate not only the representativeness of the NAFWiC, but the strength of the differences between the two levels of writing.

4.3.6 File names and annotation

Each piece of writing in the NAFWiC has an alphanumeric file name that provides four pieces of information about each file in the following order:

1. The subcorpus of the particular piece of writing. Table 4.10 presents abbreviations for each subcorpus.
2. The number of papers in this subcorpus. Currently this is two digits between 01 and 41.
3. The instructor under which the piece was written. This is a letter A through D (see Section 4.3.5).
4. The course in which the piece of writing was composed (see Section 4.3.5). Table 4.11 displays the abbreviations for each course.

Table 4.10 Subcorpus Abbreviations

<i>Subcorpus</i>	<i>Abbreviation</i>
Expository Essay	EXPO
Persuasive Essay	PERS
Personal narrative	NARR
Summary synthesis	SUMM
Rhetorical analysis	ANAL
Remembering an event	REMB
Profile	PROF
Proposing a solution	PROP
Film Analysis	FILM
Literary Research Paper	LITR
Thematic Analysis	THEM
Character Analysis	CHAR

Table 4.11 Course Abbreviations

<i>Course</i>	<i>Abbreviation</i>
English 105	105
English 1101	1101
UWP 1	1
English 113	113
English 1013	1013
English 1023	1023

For example, the file named SUMM01A105 denotes a summary synthesis paper, the first text in this subcorpus, written under Instructor A in English 105. The file named CHAR17D1023

denotes a character analysis paper, the 17th file in this subcorpus, written under Instructor D in English 1023. Most computers display files in a folder alphabetically according to the first element, which allows the NAFWiC files to appear according to subcorpus. This allows the user to easily select text types for analysis when selecting files for a concordancing program.

Each file in the NAFWiC has minimal annotations where the following information has been deleted:

<Q> long quotations in blocked format from a cited source;

<T> table;

<F> figure or image;

<U> erased URL in-text citation;

<R> all reference and bibliography sections.

Most grammatical and spelling mistakes (apostrophe misuse, *and* instead of *an*, missing “to” in the phrase “a definite threat human health”) were fixed without ado. Because these corrections do not bear direct relevance on the research question at hand (comparison of noun phrases between scholarly writing and successful first year undergraduate writing), I felt justified in simply correcting such errors. In some instances, however, a judgment call was made regarding the intention of the author. For example, in SUMM12A105, one line read “the innocents of the trees.” I believe this was intended as “the *innocence* of the trees,” and completed the change accordingly. This type of error was noted on the original file via an “Inserted Comment” function from Microsoft Word in order to track such corrections allowing for confirmation of judgments at a later date if necessary.

Additionally, each file in the NAFWiC was run through a part of speech (POS) tagger. POS tagging involves assigning grammatical categories to each word in a corpus, which can facilitate a more sophisticated analysis of a corpus. Granger and Rayson (1998), for example,

used POS tagging to compare the use of determiners, pronouns, conjunctions, prepositions, nouns and adverbs by native and non-native speakers. In this study, POS tagging is used to compare the frequency of use of lexical words (nouns, verbs, adjectives and adverbs) of successful first year undergraduate writing to that of scholarly writing.

A range of POS taggers is available, and accessing the taggers often involves providing the files to be tagged to the tagging software designer who then runs the files through the software and returns a “tagged” copy. This is the case, for example, for the POS tagger that Biber has created and is used by many of his own students at Northern Arizona University (Friginal, 2010). The tagging program used in this thesis is QTag (Mason & Tufis, 1997), a probabilistic POS tagger which works using statistical methods. QTag has been described as fairly robust and from an informal evaluation, deemed to tag with “good accuracy” (Le, 2003). Although it is highly unlikely a tagger will be 100% accurate, as noted in Hunston (2002, p. 83), global accuracy of a tagger is usually irrelevant, as “all the mistakes a tagger makes...will be clustered around words which have several possible tags...thus, a tagger with an accuracy rate of 96%, say, may be 100% accurate for many words, but only 70% accurate for some words.” Many of the issues surrounding accuracy in a tag-set (e.g. *put* as participle or past tense) are irrelevant for this thesis, whose focus is on nouns. In order to assess the accuracy of QTag for tagging nouns, however, I manually evaluated a random sample of 100 NN (common singular noun) and NNS (common plural noun) tags; the accuracy rate for this random sample was 97%, more than satisfactory for purposes of this thesis⁸ (see Appendix 3). Furthermore, in a comparison of three POS-taggers, QTag was found to be the fastest (by at least an order of magnitude compared to the slowest) and most robust when dealing with misspelt words and other ‘junk’ text (Johnson, Malhotra, &

⁸ The three instances of incorrect tagging were all related to tagging a word which could be either a noun or a verb; in these instances *values*, *needs*, and *change* were marked as nouns but were actually used as verbs.

Vamplew, 2006). QTag was also chosen due to its ease of access, as its creator is a co-supervisor of this thesis. Table 4.12 presents each POS tag applied by QTag to the NAFWiC.

Table 4.12 POS Tags

<i>POS</i>	<i>Tag</i>	<i>POS</i>	<i>Tag</i>
be	BE	Pronoun	PN
were	BEDR	Possessive	POS
was	BEDZ	personal pronoun	PP
being	BEG	possessive personal pronoun	PP\$
am	BEM	reflexive pronoun	PPX
been	BEN	adverb	RB
are	BER	comparative adverb	RBR
is	BEZ	superlative adverb	RBS
coordinating conjunction	CC	adverbial participle	RP
cardinal number	CD	symbol or formula	SYM
subordinating conjunction	CS	infinitive marker	TO
do	DO	interjection	UH
did	DOD	base verb	VB
doing	DOG	past tense verb	VBD
done	DON	progressive verb	VBG
does	DOZ	past participle verb	VBN
determiner	DT	wh- determiner	WDT
existential there	EX	wh- pronoun	WP
foreign word	FW	wh- adverb	WRB
have	HV	negative marker	XNOT
had	HVD	exclamation mark	!
having	HVG	quotation mark	“
has	HAZ	apostrophe	‘
preposition	IN	open parenthesis	(
adjective	JJ	close parenthesis)
comparative adjective	JJR	comma	,
superlative adjective	JJS	dash	-
modal auxiliary	MD	full stop	.
common singular noun	NN	ellipses	...
common plural noun	NNS	colon	:
proper singular noun	NP	semicolon	;
proper plural noun	NPS	question mark	?
ordinal number	OD	unclassified	???
pre-determiner	PDT		

4.3.7 Research and continued development

As described in this thesis, I have used the NAFWiC to examine the modified noun phrase of successful first year undergraduate writing as represented in the NAFWiC compared

with that in scholarly writing (as compiled from the data in corpus-cited references discussed in section 4.2). I have also used the NAFWiC in presentations as part of colloquia on academic writing at international TESOL conventions (Bennett, et al, 2010, 2011). Future research involving the NAFWiC may involve a comparison of North American and British academic writing, comparison of first year and upper-level undergraduate writing and/or exploration of text type, among other possibilities. In addition to a focus on research involving the NAFWiC, as mentioned earlier, many teaching implications are also available from analyses of the corpus, including but not limited to the design of corpus driven activities and/or materials development for EAP writing courses. Pedagogical applications of the research here will be discussed in Chapter 7.

The current compilation of the NAFWiC is only the beginning of corpus-based research related to freshman composition. Although the initial creation of the NAFWiC is an exciting development, the present size is not sufficient for certain statistical studies. Hopefully the NAFWiC will continue to add files from freshman composition courses. These new files could be obtained from instructors who have already contributed to this version of the NAFWiC as well as from new instructors/institutions. Adding files to the NAFWiC will allow it to grow not only in total number of tokens, but also allow it to offer a more complete range of text types. Currently, the NAFWiC is available by request via an emailed zip file; perhaps in the future the corpus could be available via web access.

4.4 Summary

To ascertain the differences between expert and novice writing as represented on the staircase model proposed in Chapter 2, this study compared the modified noun phrase in corpora representing those levels of writing; this comparison served to provide further evidence against using expert writing to inform EAP teaching, specifically in the IEP, and to

contribute to an understanding of an appropriate description of novice writing which can inform such teaching. This chapter introduced the corpora used in that comparison. The data used in constructing the modified noun phrase in scholarly (expert) writing was extracted from two general academic writing corpora representing a range of disciplines and published in two large corpus-cited references. The data used in construction of the modified noun phrase in first year undergraduate (novice) writing originated from the NAFWiC, an exemplar corpus representing the discipline-specific writing of freshman composition. This chapter presented the materials used in this study, including an overview of the corpora used in the corpus-cited references which informed the scholarly writing data and an introduction and full description of the NAFWiC. Chapter 5 discusses the methods used to extract and compare noun phrases in the NAFWiC and those from the corpus-cited references. Comparative results are presented in Chapter 6.

CHAPTER FIVE

IDENTIFYING NOUN PHRASES IN FRESHMAN AND SCHOLARLY WRITING

5.1 Introduction

As discussed in Chapter 3, corpus-based research is distinguished by its empirical analysis through both quantitative and qualitative techniques of the actual patterns of language use in a collection of authentic texts, typically using word- or context-based approaches. Context-based studies typically begin with the investigation of a register (or a context within a register), identifying the combination and application of words and categories which distinguish that context from another, often making use of statistical calculations and corpus annotations. The study reported here engages in a context-based approach, making use of frequency lists, tagging, and parsing to investigate how the noun phrase behaves differently in scholarly writing and freshman composition writing. This chapter provides an overview of the methods employed which eventually led to the identification and analysis of noun phrases in the NAFWiC (freshman writing) and corpus-cited references (scholarly writing), both of which were introduced in Chapter 4. Specifically, section 5.2 provides an overview of corpus-cited references and explains how I approached and extrapolated data from them in order to identify salient features to create a lexicogrammatical profile of scholarly writing; the profile was used to inform the comparative study between scholarly and freshman composition writing. Section 5.3 details the process of identifying, classifying, and analyzing the data in the NAFWiC in order to compare data in the scholarly writing profile to that in freshman writing. Comparative results of the study are presented in Chapter 6.

5.2 Methods Used for Scholarly Writing

5.2.1 Understanding corpus-cited references

Extensive grammar and vocabulary books which specifically cite corpus findings have been termed “corpus-cited texts” (Bennett, 2010, p. 36), or corpus-cited references. Corpus-cited texts are the products of extensive research using extremely large, generalized corpora, which provide information about how grammatical features and functions are actually used in context. Chapter 4 detailed the general academic writing corpora that were used in the corpus-cited references informing this study [the *Longman Grammar of Spoken and Written English (LGSWE)* (Biber, et al, 1999), the *Cambridge Grammar of English (CGE)* (Carter & McCarthy, 2006), and *A Comprehensive Grammar of English (CommGE)* (Leech & Svartvik, 2002)]. Corpus-cited texts are useful for both teachers and researchers. Teachers can consult corpus-cited references to facilitate the creation of lessons and/or supplement published teaching materials to better inform teaching; specifically, verifying information from published textbooks with that reported in corpus-cited references can make teaching materials more effective¹ (Bennett, 2010). However, teachers should approach corpus-cited references with caution, as their thousands of pages and seemingly unending bits of information have a tendency to be overwhelming. Three strategies for dealing with corpus-cited texts include making good use of the conceptual as well as lexical indices, knowing what you want to research before you open the reference, and skimming the information in the texts (Bennett, 2010, p. 36).

As an example of working with corpus-cited references to make teaching materials more effective, Bennett (2010, p. 40-43) discusses working with adverb clauses of time while teaching from *Understanding and Using English Grammar (UEEG)* (Azar, 2002), an

¹ Benefits of using “corpus-informed” materials are discussed more in Chapter 7.

extremely popular grammar book in North American settings. Table 5.1 presents information about adverb clauses of time taken from chapter 5 of *UUEG*, specifically charts 5.1 and 5.2 (Azar, 2002, p. 70-72). The information presented in the charts almost certainly looks reliable from the point of view of a teacher’s intuition about adverb clauses of time, and students have likely never had trouble with the presented information.

Table 5.1 Key ideas about adverb clauses of time (Azar, 2002, p. 70-72)

<p><i>When the phone rang</i>, the baby woke up. The baby woke up <i>when the phone rang</i>.</p>	<p>These examples have the same meaning. An adverb clause can come in front of the main clause or follow the main clause.</p> <p>Notice that a comma is used to separate the two clauses when the adverb clause comes first.</p>	
<p>Common adverb clauses to show time relationships</p>		
<p><i>after*</i>, <i>before*</i>, <i>when</i>, <i>while/as</i>, <i>by the time</i>, <i>since</i>, <i>until/till</i>, <i>as soon as/once</i>, <i>as long as/so long as</i>, <i>whenever/every time</i>, <i>the first/last/next time</i></p>	<p>* <i>After</i> and <i>before</i> are commonly used in the following expressions:</p>	
	<p><i>shortly after</i> <i>a short time after</i> <i>a little while after</i> <i>not long after</i> <i>soon after</i></p>	<p><i>shortly before</i> <i>a short time before</i> <i>a little while before</i> <i>not long before</i></p>

However, a review of the information concerning adverb clauses of time provided in a corpus-cited reference compared with that published in the textbook provides enlightening information for the teacher. Reviewing the tables from the *LGSWE* (Biber et al., 1999), reproduced in Figures 5.1-5.4 below, leads to the discovery of three core pieces of information which should be included in any lesson on adverbs of time: adverb clauses of time mostly come at the end of a sentence, especially in writing; *when* is by far the most common subordinator used in adverb clauses of time, in all registers; *since* and *as* are not normally used as adverb clauses of time in academic prose, but are especially common as adverb clauses of time in conversation (Bennett, 2010, p. 42). These pieces of information specifically differ from the information in Table 5.1 in two critical areas. Firstly, the issue of

5.2.2 Approaching corpus-cited references

5.2.2.1 A Systematic Approach

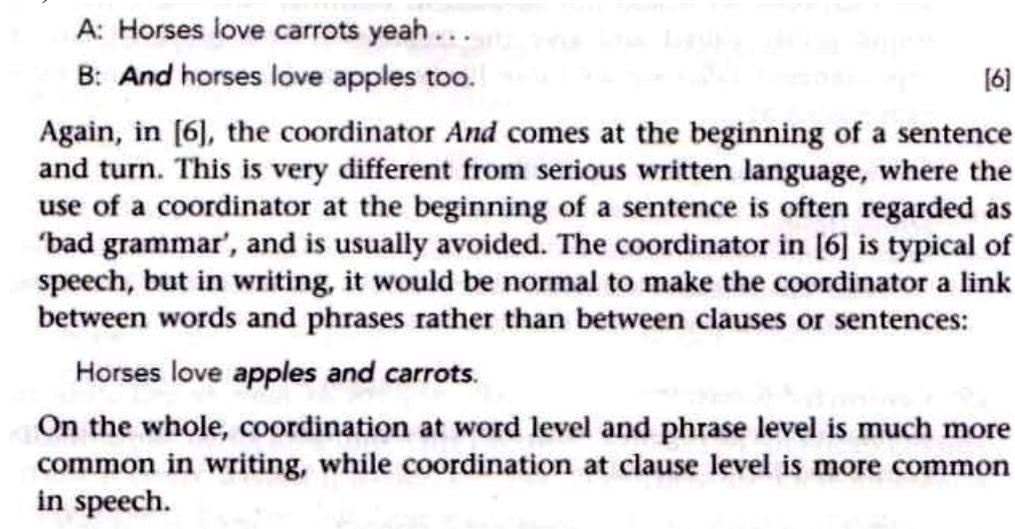
As discussed above, working with corpus-cited references may initially feel intimidating; the sheer amount of data and findings contained in the texts literally covers thousands of pages. A systematic approach to the references is the only option for successfully navigating the treasure trove of data. In order to identify linguistic features and functions from the corpus-cited references for compilation of a lexicogrammatical profile of scholarly writing, such an approach was used.

Perhaps it goes without saying, the Table of Contents is the obvious starting place for approaching any book, including these corpus-cited references. Because I knew my focus was on scholarly writing (termed “academic prose” in the *LGSWE*; “academic language, written style” in the *CGE*; and “formal or serious writing” in the *CommGE*), I first scrutinized the contents of each corpus-cited reference for mention of the academic register.

Surprisingly, the 17 page “Contents” of the *LGSWE* provides little information about scholarly writing. A description of scholarly writing as represented in the *LGSWE* (including an overview of the academic subcorpus discussed in Chapter 4) is part of the Introduction, and although register is mentioned in various subsections of chapters (e.g. 2.2.1.2 *TTR across the registers*), only chapters 13 and 14 (“Lexical expressions in speech and writing” and “The grammar of conversation,” respectively) focus on register; the other 11 chapters in the reference are organized around key word classes. Access to register information, other than introductory, is not provided via the table of contents in the *LGSWE*. Instead, the *LGSWE* has to be examined chapter by chapter to note reference to scholarly writing; thus, the chapter structure, discussed below, is critical to accessing findings via register in this reference. For the *CGE*, however, although the “Contents” presents chapters organized around key word

classes, chapter 5 specifically deals with “Grammar and academic English.” The chapter “focuses on items and structures which are common in academic language and which characterise it” (Carter & McCarthy, 2006, p. 267). Thus, this entire chapter from *CGE* became a focus point for collecting data and findings for the scholarly writing profile. The chapters in *CommGE* are organized according to grammatical function. In this sense, the *CommGE* is similar to the *LGSWE* in that reference to the register is made within each chapter. Thus, the *CommGE* was also examined chapter by chapter to note reference to formal writing. Figure 5.5 demonstrates such a reference to “serious written language” regarding coordination in the *CommGE*. As noted in Chapter 4, the *CommGE* was useful in confirming what was reported in the *LGSWE* and *CGE*.

Figure 5.5 Reference to formal writing in relation to coordination (Leech & Svartvik, 2002, p. 15)



In the *LGSWE*, closer examination of each chapter and section reveals a clear pattern for the presentation of data via register. Many of the subsections in each chapter present “Corpus Findings” as a series of bullet points which review the most important findings from the data; these points are accompanied by graphs and tables presenting the data via register followed by a prose explanation of the data in the “Discussion of Findings” section. Figures

5.6 and 5.7 display the Corpus Findings, graphs and tables, and the Discussion of Findings from the *LGSWE* reporting on pronouns versus full noun phrases (Biber, et al., 1999, p. 235-236). The graph in Figure 5.6 clearly shows the preference of scholarly writing for nouns (versus pronouns), with approximately 300,000 nouns per million words as compared to fewer than 50,000 pronouns per million words. The table in Figure 5.7 further demonstrates the preference of nouns in scholarly writing (versus pronouns), but also shows that when pronouns are used in scholarly writing, they occupy the subject position. For clear and instant information concerning the behavior of any word class in a given register, the graphs and tables provided in the Corpus Findings section of each chapter are the key starting point; when compiling information to create the profile of scholarly writing, these tools were, indeed, the first point of contact, especially relating to numerical frequencies and results. The majority of frequencies reported below in context with the scholarly writing profile, as well as those for the noun phrase in Chapter 6, were obtained from these particular chapter sections of the *LGSWE*.

Figure 5.6 Pronouns v Full Noun Phrases: Corpus Findings, Discussion of Findings, and Graph (Biber, et al., 1999, p. 235)

Pronouns v. full noun phrases

The distribution of nouns and pronouns varies greatly depending upon register (2.3.5, 2.4.14). It further turns out that the use of pronouns v. full noun phrases varies in relation to syntactic role.

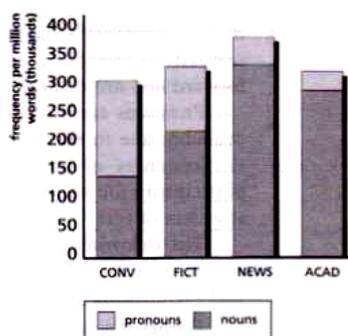
CORPUS FINDINGS 3.16

- Pronouns are slightly more common than nouns in conversation.
- At the other extreme, nouns are many times more common than pronouns in news and academic prose.
- The noun-pronoun ratio varies greatly depending upon syntactic role.
 - The relative frequency of nouns is much higher in object position and as a complement or object of a preposition than in subject position.

DISCUSSION OF FINDINGS

As illustrated in 4.1.1, there are important differences in the reliance on nouns v. pronouns across registers. In conversation, the shared situation and personal involvement of the participants result in a dense use of pronouns. In contrast, the informational purposes of news and academic prose result in a much more frequent use of nouns, and proportionally many fewer pronouns. (See also 2.3.5, 2.4.14.)

Figure 4.1
Distribution of nouns v. pronouns across registers



provided in the *CGE*; rather, phrases like “relatively common” or “extremely frequent” are used. The *CGE* does, however, provide more detail about functions and use of the selected features of scholarly writing outlined in the chapter. Figure 5.8 displays information about the function of noun phrases in academic English. As noted in the figure, noun phrases allow for definition and specification, useful in academic English. The majority of functions reported below in context with the scholarly writing profile were obtained from the *CGE* following this approach.

Figure 5.8 Discussion of Function of Noun Phrases (Carter & McCarthy., 2006, p. 269)

Post-head elements: defining and specifying 141d

Postmodified and complemented noun phrases are extremely frequent in academic English because of the frequent need for definition and specification. All the types of postmodification and complementation described in **The noun phrase** are used; → 171-172.

Prepositional phrases are very frequent, and may consist of several occurring together:

(noun phrases in green, prepositions beginning post-head elements in bold type)

From *their differing positions* **within the family**, men and women separately weighed the potential benefits and **risks of migration**.

5.2.2.2 Salient Features of Scholarly Writing

It is widely accepted that users of English exploit the language according to context. Biber (2006, p. 6) explains this concept by stating that “linguists have come to recognize that language characteristics differ dramatically from one register to the next...”, and as studies such as those discussed in Chapter 3, section 3.5 have shown, from one discipline to another. Carter & McCarthy (2006), however, seem to caution that the exploitation of English is never so extreme as to merit unique structures for specific contexts; in regards to scholarly writing, they state that “most of the grammar of academic English is shared with that of English as a whole, and there are no special structures which are unique to academic English and never

found elsewhere” (Carter & McCarthy, 2006, p. 267). While Biber (2006) also agrees that “there are few general linguistic features that are uniquely characteristic of academic prose... larger set of features—such as nouns and prepositional phrases—occur to some extent in every register” (p. 18), he understands that certain features can be identified as “academic” because they are especially common in scholarly writing. As discussed in Chapters 3 and 4, however, even these “academic” features often cluster in various manners depending on text type, genre, and discipline. The compilation of a lexico-grammatical profile exhibiting this “larger set of features” which characterizes scholarly writing was the first step toward confirming differences in scholarly and first year undergraduate writing in order to provide an appropriate instructional model for teaching EAP writing in the IEP.

Using the systematic approach described in section 5.2.2.1 to extensively review the three corpus-cited references presented in Chapter 4, I compiled a full description of linguistic features and functions in scholarly writing as reported in the references. A complete list of linguistic features of scholarly writing identified from this review and categorized by part of speech is provided in Appendix 4; a list of functions and their role in scholarly writing is provided in Appendix 5. Close examination of these overall findings reveals those features and functions which are both positively and negatively salient in scholarly writing; that is, these features and/or functions were reported as occurring overwhelmingly in scholarly writing or occurring more often in scholarly writing than other registers (positive) or reported as rarely occurring in scholarly writing or occurring more often in other registers than in scholarly writing (negative). A list of negative salient features and functions in scholarly writing is provided in Appendix 6. Positive salient features and functions of scholarly writing are presented here. This list is particularly important as it reveals the overall structures and functions of language which characterize scholarly writing.

- 60% of lexical words in scholarly writing are nouns (Biber et al, 1999, p. 65)
Plural nouns are used 3-4 times more in scholarly writing than conversation (Biber, et al., 1999, p. 291).+²
- Definite noun phrases are used in anaphoric expressions in scholarly writing whereas other registers use pronouns in anaphoric expressions (Biber, et al., 1999, p. 266).*
- Cataphoric reference is more common in scholarly writing than other registers (Biber, et al., 1999, p. 266).+
- Nominalization is much more common in scholarly writing than other registers, especially *-tion* and *-ity* (Biber, et al., 1999, p. 322).+
- Preposition+*which* relativizers are only common in scholarly writing (Biber, et al., 1999, p. 625). *of-which* and *whose* are used equally in scholarly writing whereas other registers use *whose* most (Biber, et al., 1999, p. 618).@*
- The comparative adjective is used more in scholarly writing than in any other register; it is used three times more in scholarly writing than conversation; conversely, superlatives are rare in scholarly writing (Biber, et al., 1999, p. 524).+
- Scholarly writing makes use of degree modifiers whereas conversation prefers amplifiers (Biber, et al., 1999, p. 565).*
- In written English, we often use non-finite and verbless clauses as adverbials and modifiers. Such constructions are unlike in speaking where finite clauses are preferred (Leech & Svartvik, 2002, p. 15-16).*
- Adverbial *ing* and *ed* clauses are particularly favored in formal written styles (Leech & Svartvik, 2002, p. 203).#
- Subjunctive is quite common in written English (Leech & Svartvik, 2002, p. 396-397).#
- Abstract nouns and *it*-begin sentences characterize impersonal language of formal written texts (Leech & Svartvik, 2002, p. 33).#
- Coordination at word and phrase level is more common in writing whereas coordination at clause level is more common in speech (Leech & Svartvik, 2002, p. 15).*
- Latin abbreviations *i.e. viz, e.g.* are generally only common in formal written English (Leech & Svartvik, 2002, p. 189).@

² The symbols following each statement are explained at the end of the list.

- Rhetorical questions and questions which are immediately answered are frequent in academic contexts (Carter & McCarthy, 2006, p. 288).#
- In general, academic writing displays quite complex sentence patterns, including frequent use of all types of subordination (Carter & McCarthy, 2006, p. 288).#
- Non-finite subordinate phrases and ellipted subordinate clauses are frequent in academic writing (Carter & McCarthy, 2006, p. 289).#
- Single adverbials are used differently in scholarly writing than other registers:*
 - 65% of uses of *as* are for manner in scholarly writing, but for time and reason in fiction/news (Biber, et al., 1999, p. 847);
 - 95% of uses of *since* are for reason in scholarly writing, but for time in other registers (Biber, et al., 1999, p. 848);
 - 80% of uses of *while* are for concession in scholarly writing, but 100% for time in conversation (Biber, et al., 1999, p. 849).
- In scholarly writing, *though* is used 95% of the time as a subordinator, but 90% of the time as a linking adverbial in conversation (Biber, et al., 1999, p. 850).*
- Existential *there* is used over five times more than locative *there* in scholarly writing. While scholarly writing has about the same frequency of use of existential *there* as other registers, it uses locative *there* so much less (Biber, et al., 1999, p. 948).*
- Use of *and/or* is not very common in scholarly writing, but it is rarely found elsewhere and is, therefore, basically unique to scholarly writing (Biber, et al., 1999, p. 81).@
- The use of dual gender reference (*he or she*) is unique to scholarly writing (Biber, et al., 1999, p. 317).@
- Coordination tags *or so*, *and so on*, and *etc.* are used commonly only in scholarly writing (Biber, et al., 1999, p. 116).+
- The overwhelming use of classifiers (nouns and attributive adjectives) is unique to scholarly writing, particularly relational and topical/miscellaneous classifiers (Biber, et al., 1999, p. 506-511).+
- Scholarly writing uses the *-ly* suffix for adverbs more than other registers (Biber, et al., 1999, p. 540).+
- *That/those + of*-phrase is common only in scholarly writing (Biber, et al., 1999, p. 308).+

- Linking adverbials only account for less than 10% of adverbials in scholarly writing, but this is a greater use than any other register (Biber, et al., 1999, p. 765);@
- Verbs used in post predicate *that* clauses in scholarly writing are used to indicate degree of certainty with reported information. Other registers use mental verbs to express stance of a person, which is rare in scholarly writing (Biber, et al., 1999, p. 666-667).*
- Although only 25% of verbs in scholarly writing are passive, scholarly writing uses the passive voice more than any other register (Biber, et al., 1999, p. 476).@
- Post modifiers in a noun phrase with short passives are more than two times as common in scholarly writing than other registers (Biber, et al., 1999, p. 938).+
- Long passives rarely occur in scholarly writing; however, they do not occur in any other register at all, making them unique to scholarly writing (Biber, et al., 1999, p. 940).@
- Idiomatical phrases with *have, make, and take* are more common in scholarly writing than conversation (Biber, et al., 1999, p. 1028).*
- The adverb *approximately* is used primarily in scholarly writing (Biber, et al., 1999, p. 113).+
- Post modified and complemented noun phrases are extremely frequent in academic English (Carter & McCarthy, 2006, p. 269).#
- Expressions such as *in my opinion/I would suggest, argue/it is reasonable* are preferred in academic writing (Carter & McCarthy, 2006, p. 278).#

These features and functions which characterize scholarly writing can be divided into four categories:

- those which are not necessarily common in scholarly writing but are typically only found in scholarly writing (and not in other registers). For example, linking adverbials account for less than 10% of adverbials in scholarly writing, but this is a greater use than in any other register (Biber, et al., 1999, p. 765). (These features and functions are marked with a @; there are 7 such features.).

- those features with frequency or function which is in contrast to that of other registers. For example, scholarly writing makes use of degree modifiers whereas conversation prefers amplifiers (Biber, et al., 1999, p. 565). (These features and functions are marked with a *; there are 10 such features.).
- those that are common in scholarly writing (without comparison to other registers). For example, post modified and complemented noun phrases are extremely frequent in academic English (Carter & McCarthy, 2006, p. 269). (These features and functions are marked with a #; there are 8 such features.)
- those that are more common in scholarly writing than other registers. For example, scholarly writing uses the *-ly* suffix for adverbs more than other registers (Biber, et al., 1999, p. 540). (These features and functions are marked with a +; there are 7 such features.)

The features and functions in the list are rather evenly divided into the four categories, though as would be expected, the smallest category is that which contains features or functions that are found to any degree of significance only in scholarly writing.

An immediate question arises from a list such as this: Is it useful? The simple answer is yes, such a list is useful. These characteristics would be obvious teaching points for writing instruction which would target an expert writing model for a number of reasons; they are, after all, those features and functions that characterize scholarly writing. Firstly, the features and functions on this list could resolve the “what to teach” issue as discussed in Chapter 3 section 3.5.7; especially those features which are noted as common in scholarly writing (those marked with a + or #). Secondly, of particular importance would be those features with frequencies or functions which are in contrast to those of other registers (those marked with a *). Given that novice (levels 1 and 2) and even developing (level 3) writers have a tendency

to apply casual spoken language to formal written work (Hinkel, 2002a, 2003; Shaw & Liu, 1998), it would be important to bring these variant uses to students' attention. Those features and functions which are rarely found outside of academic prose would also be particularly useful for teaching (those marked with a @); it is unlikely students will encounter these features in their studies otherwise.

While such a list can be said to be very useful to inform instruction in the ways just mentioned, two concerns do arise. Realizing such statements as “only 25% of verbs in scholarly writing are passive” is one issue, both in terms of ability and usefulness, that has plagued the applicability of corpus findings to language teaching materials (Frankenberg-Garcia, 2006). This issue is addressed more fully in Chapter 7 regarding pedagogical applications. The other issue arising from use of this list involves its actual implementation to inform writing instruction. The third sentence in the paragraph above claimed that “These characteristics would be obvious teaching points for writing instruction which would target an expert writing model for a number of reasons...” The caveat here is *which would target an expert writing model*. Given the theoretical position informing the staircase model proposed in Chapter 2, namely the ZPD and input hypothesis, only those learners at level 3 on the staircase, graduate writing, would use an expert writing model. And while the statements on the list may serve to inform writing instruction at that level, as discussed in Chapter 3, graduate level writing should also be considered in terms of specific disciplines. For teaching EAP writing in the IEP there is actually a danger of having such a list available; instructors and materials writers would want to use it to inform instruction, and as argued in this thesis, expert writing is a poor model for novice writers. These concerns notwithstanding, this list can be useful to inform writing research, research at all levels of writing proficiency. Having a profile of level 4 writing provides points of comparison for the other levels on the staircase

(or across disciplines on each level), thus revealing how writers' use of the "larger set of features" develops as they ascend the staircase. The present research was derived from this idea.

5.2.3 *The noun phrase in scholarly writing*

The list of positive salient features and functions of scholarly writing presented above characterizes scholarly writing. Closer investigation of each point on the list gives further details about the structures and functions and their roles in scholarly writing. Consider the first point on the list, "60% of lexical words in scholarly writing are nouns" (Biber et al, 1999, p. 65). Looking into nouns in scholarly writing, given their prevalence, leads to other noteworthy aspects. Figure 5.9 displays the distribution of lexical word classes across the core registers of the *LGSWE*; Figure 5.10 specifically displays the distribution of modification of nouns in scholarly writing. Clearly, nouns, and modified ones, are favored in scholarly writing. In fact, the most salient linguistic feature of scholarly writing is its extensive use of nouns and noun phrases. As shown in Figures 5.9 and 5.10, 300,000 of 500,000, that is 60%, of all content words in scholarly writing are nouns (Biber et al, 1999, p. 65); of those nouns, 60% are modified (Biber et al, 1999, p. 578). Conversely approximately 150,000 of 350,000, that is 40%, of content words in conversation are nouns (Biber et al, 1999, p. 65), with little more than 15% of noun phrases in conversation being modified (Biber et al, 1999, p. 578).

Figure 5.9 Distribution of Lexical Word Classes across Registers (Biber et al, 1999, p. 65)

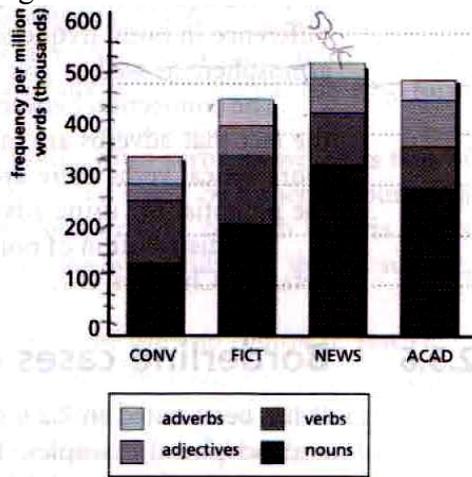
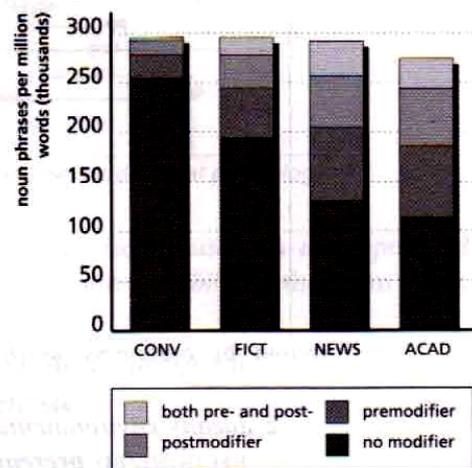


Figure 5.10 Distribution of Noun Phrases³ with Pre-Modifiers and Post-Modifiers (Biber et al, 1999, p. 578)



Two elements are required for a noun phrase: a determiner and a noun. Figure 5.11 displays the use of definite and indefinite articles in different syntactic roles across the written registers of the *LGSWE*. Scholarly writing makes more frequent use of the definite article overall, but marks a clear preference for *the* in subject position. Figure 5.12 displays the distribution of definite determiners across registers. Again, scholarly writing has an overwhelming preference for the definite article. As shown in Figures 5.11 and 5.12, the

³ Common nouns are the most frequent type of noun phrase head, and they are also the most productive head type occurring with both pre- and post-modifiers (Biber, et al, 1999, p. 581). As such, all figures and findings reported here are based specifically on the “common head noun.”

and post- modifier usage and types thereof as well as frequency of the definite article in scholarly writing and successful first year composition writing as represented in the NAFWiC. The core of the comparison between the two registers centers on modification patterns.

5.2.4 Summary

The general academic writing corpora discussed in Chapter 4 informed the corpus-cited references used in this research. Section 5.2 discussed how corpus-cited references can be used by both instructors and researchers and specifically detailed the methods I used to approach the corpus-cited references to obtain data for use in the development of a profile of scholarly writing. That profile led to and informed a comparison study of scholarly and freshman composition writing, specifically the use of modified noun phrases, which aimed to confirm differences in those two levels of writing in order to provide an appropriate instructional model for teaching EAP writing in the IEP.

5.3 Methods Used for the NAFWiC

5.3.1 Overview

As discussed in section 5.1, context-based studies in corpus linguistics typically begin with the investigation of a register (or a context within a register, such as freshman composition within academic writing) identifying the combination and application of words and categories which distinguish that context from another, often making use of frequency and corpus annotations involving both qualitative and quantitative methods. The study reported here engages in a context-based approach comparing expert and novice writing to establish differences between the two levels of writing. Section 5.3 details the methods used in the NAFWiC to engage in that comparison. Here I provide an overview of the processes involved in the corpus annotations and frequency count.

Firstly, the NAFWiC was annotated for part of speech (POS) using the QTag program (see Chapter 4 section 4.3.6 for an overview of the POS tags). Following this, TextSTAT (Hüning, 2008), a freely available concordancer, was used to establish frequency counts of the POS tags in order to determine the distribution of lexical word classes in the NAFWiC; this is the first point of comparison between scholarly and freshman composition writing. Thirdly, files in the NAFWiC were run through a noun phrase extractor (Mason, 2010) to allow for the identification of noun phrases in the NAFWiC. However, the noun phrase extractor was not able to differentiate between top-level head nouns and others, such as those as object of a preposition position. Because the noun phrase extractor could not automatically make such a distinction, the noun phrase identification in the NAFWiC required manual analysis, whereby I read through each of the 246 files manually identifying each top-level noun and, where applicable, its accompanying phrasal parts, creating a list of noun phrases in each file. Section 5.3.2 details the processes involved in identifying and classifying noun phrases in the NAFWiC, and section 5.3.3 discusses the methods employed for quantitative analysis of the noun phrase.

5.3.2 Identifying noun phrases

5.3.2.1 Modification Patterns

After each top-level noun phrase had been identified in the NAFWiC, the list of phrases from each file was converted to table format in Microsoft Word⁴ using the “convert text to table” tool in order to allow for easier quantitative analysis (discussed in 5.3.3 below). The tables were then copied into Microsoft Excel. Each subcorpus was saved as a separate Excel workbook, with the list of noun phrases from each file saved in a separate worksheet. For example, following this process, the LITR subcorpus contains six worksheets: five file

⁴ Microsoft Word and Excel 2002 were used to format and analyze the data.

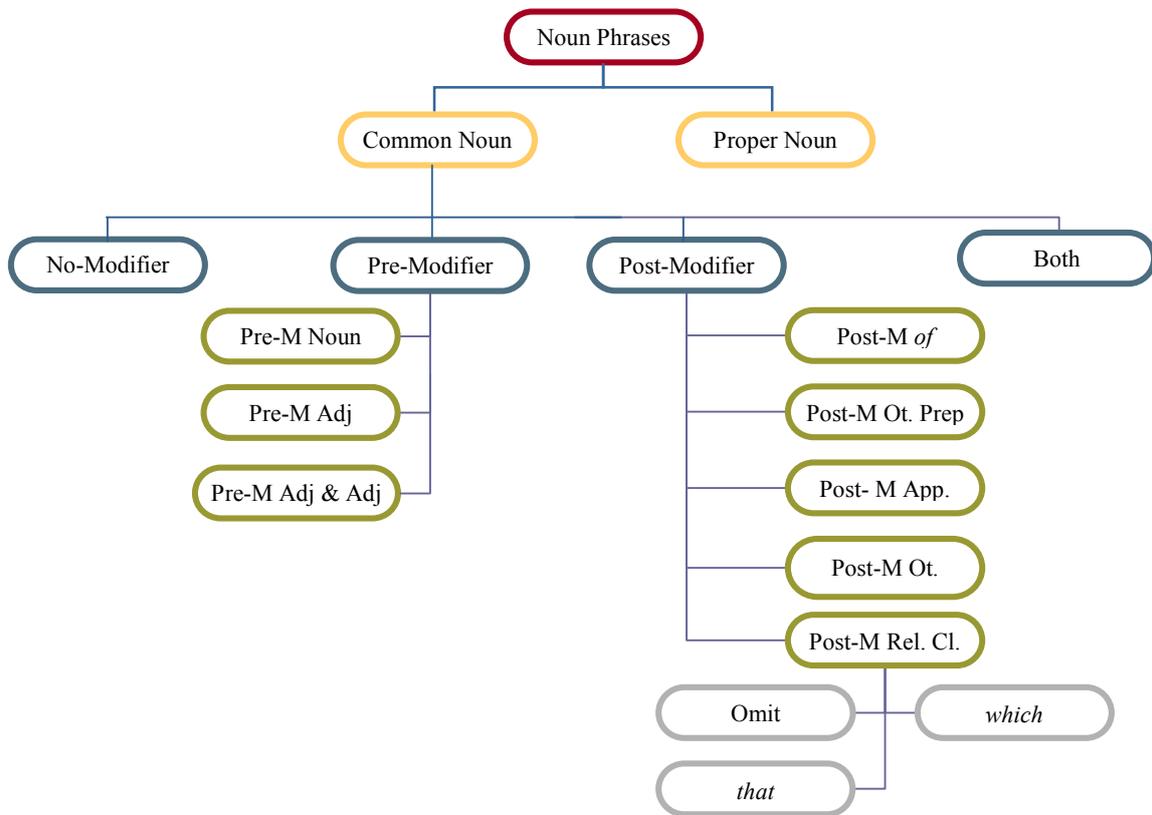
worksheets converted from the list of the noun phrases in each file plus a summary worksheet displaying the quantitative data from each file for the LITR subcorpus as a whole. Each of these six worksheets for the LITR subcorpus is included in Appendix 7. An additional workbook was created to display the data for the NAFWiC as a whole; this workbook contains only one summary worksheet, but uses Excel formulas to automatically pull the data from the summary worksheets in the other 12 workbooks. The final result of the conversion process is 13 Excel workbooks (one for each individual subcorpus and one for the NAFWiC as a whole) containing 259 worksheets [one for each individual file (246 file worksheets), each individual subcorpus (12 summary worksheets), and the summary worksheet for the NAFWiC as a whole]. The 13 Excel workbooks can be found on CD-Rom in Appendix 8.

After converting the list of noun phrases from each file into its own worksheet, each noun phrase was then classified from one to four times (See Figure 5.13). Firstly, nouns were classified as either common or proper; as mentioned in section 5.2.3, figures and findings in scholarly writing were based only on the common noun; therefore, only common nouns needed to be included in the NAFWiC analysis. Common nouns were then further classified as having no modifier, a pre-modifier, a post-modifier, or both a pre- and post-modifier. Those with pre- and post-modifiers were further classified. Pre-modifiers were classified into three categories: noun, adjective, adjective + *and* + adjective⁵. Post-modifiers were classified into five categories: *of*-phrase, other prepositional phrase, appositive, relative clause, and other; relative clause post-modifiers were categorized one additional time as omit relativizer, *which* relativizer, *that* relativizer⁶. As discussed below, a raw frequency count of each feature was established, converted to a percentage, and later displayed in a schematic representation (discussed in Chapter 6).

⁵ This category was chosen mainly because of its marked function in scholarly writing.

⁶ The decision to further classify relative clauses only by the three relativizers omit, *which* and *that* arose from their frequency in scholarly writing, especially as compared to other registers.

Figure 5.13 Classification of Noun Phrases in the NAFWiC



As discussed in section 5.3.1 above, all noun phrases in the NAFWiC were manually identified. After being copied to the Excel file worksheet, the noun phrases were then manually classified into the categories above; each category is represented by one column in the file worksheets. See Table 5.2. Upon its classification, each noun phrase was placed in the appropriate column. This allowed for quantitative analysis discussed below. It should be noted that the column name on each file worksheet is also what appears on the front summary worksheet in each workbook. Again, see Appendix 7 for the worksheets from the LITR corpus.

Table 5.2 Column Names for the File Worksheet

<i>Classification Level</i>	<i>Column Name</i>	<i>Classification Level</i>	<i>Column Name</i>
Proper Noun	Proper Noun	Post-Modifier of	Post-M of
No modifier	No modifier	Post-Modifier Other Preposition	Post-M ot. Prep
Pre-modifier	Pre-modifier	Post-Modifier Appositive	Post-M app

Post-modifier	Post-modifier	Post-Modifier Other	Post-Mod Other
Both	Both	Post-Modifier Relative Clause	Post-Modifier Rel Cl
Pre-modifier Noun	Pre-M Noun	Omit Relativizer	Omit Relativizer
Pre-modifier Adjective	Pre-M Adj	<i>which</i> Relativizer	<i>which</i> Relativizer
Pre-modifier Adjective + <i>and</i> + Adjective	Pre-M Adj and Adj	<i>that</i> Relativizer	<i>that</i> Relativizer

5.3.2.2 Definite Article

To determine the frequency of the definite article in the subject position, two methods were used⁷. In the first method, after the list of noun phrases had been copied to the file worksheets, I manually found the letter sequence “t-h-e” in each worksheet to determine if the letters were a single word or part of a larger word; if the latter, I changed “t-h-e” within the larger word to “XXX” (e.g. “the father” was manually changed to “the faXXXr.” During this process, however, it was discovered that the manual “find and replace” function in Word could be automated to search for *the* as a single word if done prior to copying the noun phrases into the Excel file worksheet. The second method followed this discovery; all instances of *the* were replaced with “XXX” using Word’s find and replace function before the list was copied to Excel (e.g. “the father” became “XXX father”) thus reducing a manually intensive process into a simple “find and replace all” automatic process. Because of the change in method, “XXX” represents the letter sequence “t-h-e” within a larger word in some subcorpora but the single word *the* in other subcorpora depending on if the files were analyzed before or after the “short cut” was devised. Following these methods, I created Excel formulas to search for and count all instances of either “XXX” or *the*, depending on the representation of *the* in that subcorpus. Hence, in separate subcorpora, depending on when in

⁷ Unlike Word, Excel does not have a “count whole words only” option, so any instance of the letters t-h-e were initially counted, whether a whole word or not. A work-around (explained above) was then devised.

the process they were analyzed, “XXX” may represent a whole word *the* (XXX father), or it may represent “t-h-e” within a larger word (the faXXXr).

5.3.3 Analyzing noun phrases

5.3.3.1 Introduction

As discussed above, one Excel workbook was created for each subcorpus; the workbook contains one file worksheet for each of the files in the subcorpus as well as one summary worksheet which displays the quantitative data for the subcorpus as a whole. After the noun phrases were identified and classified as described above, Excel formulas were created to calculate raw frequency and percentages for each category as well as to check the accuracy of the manual classification. This is the quantitative data appearing on the summary worksheet for each subcorpus. The summary worksheet contains 35 columns for data calculations (one for raw frequency of total noun phrases in the subcorpus and one for raw frequency and one for percentage for each of the 16 categories in Table 5.2 as well as *the*) and four columns for data quality control. See Appendix 7 for the LITR summary worksheet or Appendix 8 on CD-Rom for all Excel workbooks. The sections below explain the contents of each column and the calculations completed there.

5.3.3.2 N.P. Raw

The total number of noun phrases found in the workbook. Calculated by counting the number of non-blank cells in the No Modifier, Pre Modifier, Post-Modifier, and Both columns.

5.3.3.3 No Modifier: Raw

The number of common head noun phrases with no modifier. Calculated by counting the number of non-blank cells in the No Modifier column. Examples include *action*, *readers*, *people*, *a defect*, and *sanity*.

5.3.3.4 No Modifier: %

The percentage of common head noun phrases in the text that do not contain a modifier. Calculated by dividing the raw number of noun phrases with no modifiers by the raw number of total noun phrases in the text (*No modifier: Raw # / N.P. Raw #*).

5.3.3.5 Pre-Modifier: Raw

The number of common head noun phrases with a modifier preceding the noun. Calculated by counting the number of non-blank cells in the Pre-Modifier column. Examples include *fair game, hidden wires, individual sanity, no definite black outlines, and this act*.

5.3.3.6 Pre-Modifier: %

The percentage of common head noun phrases in the text with a pre-modifier. Calculated by dividing the raw number of noun phrases with a pre-modifier by the raw number of total noun phrases in the text (*Pre-modifier: Raw # / N.P. Raw #*).

5.3.3.7 Post-Modifier: Raw

The number of common head noun phrases with a modifier following the noun. Calculated by counting the number of non-blank cells in the Post-Modifier column. Examples include *the ascertains that, the memories and love for, the character Desdemona, people who and the way [that]*.

5.3.3.8 Post-Modifier: %

The percentage of common head noun phrases in the text with a post-modifier. Calculated by dividing the raw number of noun phrases with a post-modifier by the raw number of total noun phrases in the text (*Post modifier: Raw # / N.P. Raw #*).

5.3.3.9 Both: Raw

The number of common head noun phrases with a modifier both following and preceding the noun. Calculated by counting the number of non-blank cells in the Both

column. Examples include *their voices from*, *a good time with*, *an electron cloud of*, *the only home that*, and *the Miller family from*.

5.3.3.10 Both: %

The percentage of common head noun phrases in the text with both a pre- and post-modifier. Calculated by dividing the raw number of noun phrases with both a pre- and post-modifier by the raw number of total noun phrases in the text (*Both: Raw # / N.P. Raw #*).

5.3.3.11 *the*: Raw

The number of subject noun phrases that contain the definite determiner *the*. Calculated by counting the number of *thes* or “XXX” in the worksheet (depending on the method of identifying *the*; see section 5.3.2.2 above).

5.3.3.12 *the*: %

The percentage of subject noun phrases containing the definite determiner *the*. Calculated by dividing the raw number of noun phrases with *the* by the raw number of total noun phrases in the text (*the: Raw # / N.P. Raw #*).

5.3.3.13 Proper Noun: Raw

The number of noun phrases that contain a proper noun as a head noun. Calculated by counting the number of non-blank cells in the Proper Noun column. [Although proper nouns as head nouns were not counted for the overall total number of noun phrases and were not included in the data analyzed for purposes of this thesis, for example *the Millers who* was not counted as a post-modified noun phrase or *the United States* was not counted as noun phrase with no modifier, proper nouns do appear in the data as they served as pre- and/or post modifiers (e.g. *the Miller family*, *novelist Salman Rushdie*).]

5.3.3.14 Proper Noun: Factor

The number of noun phrases that contain a proper head noun compared to the number of noun phrases that contain a common head noun. Calculated by dividing the raw number of common head noun phrases by the raw number of proper head noun phrases in the texts (*Proper Noun: N.P. Raw # / Raw #*). This factor simply allows for the comparison of the number of proper head nouns with the analyzed number of common head nouns.

5.3.3.15 Pre-M Noun: Raw

The number of common head noun phrases that employ nouns as pre-modifiers. Calculated by counting the number of non-blank cells in the Pre-M Noun column. Examples include *health insurance, factory inspectors, group therapy, story line, and family doctors*.

5.3.3.16 Pre-M Noun: %

The percentage of common head noun phrases with a noun pre-modifier. Calculated by dividing the raw number of noun phrases with a noun pre-modifier by the raw number of total noun phrases with a pre-modifier plus both a pre- and post-modifier (*Pre-M Noun: Raw # / Pre-modifier: Raw # + Both: Raw #*).

5.3.3.17 Pre-M Adj: Raw

The number of common head noun phrases with pre-modifiers that are adjectives. Calculated by counting the number of non-blank cells in the Pre-M Adj column. Examples include *emotional pain, most people, 300,000 deaths, the beaded curtain, and a better place to*.

5.3.3.18 Pre-M Adj: %

The percentage of common head noun phrases with an adjective pre-modifier. Calculated by dividing the raw number of noun phrases which utilize an adjective as a pre-

modifier by the raw number of total noun phrases with a pre-modifier plus both a pre- and post-modifier (*Pre-M Adj: Raw # / Pre-modifier: Raw # + Both: Raw #*).

5.3.3.19 Pre-M Adj *and* Adj: Raw

The number of common head noun phrases with the pre-modifier adjective + *and* + adjective + noun. Calculated by counting the number of non-blank cells in the Pre-M Adj and Adj column. Examples include *physical and emotional stressors, a populated and ethically diverse state, current and former employees from, petty and greedy people, and free and easy parking facility*. When a pre-modified noun phrase fit the adjective + *and* + adjective pattern as well as was a pre-modifier noun (e.g. *health and mind problems*), the adjective + *and* + adjective pattern took precedence, and the noun phrase was categorized in the adjective + *and* + adjective column only.

5.3.3.20 Pre-M Adj *and* Adj: %

The percentage of common head noun phrases with an adjective + *and* + adjective pre-modifier. Calculated by dividing the raw number of noun phrases with adjective + *and* + adjective pre-modifier by the raw number of total noun phrases with a pre-modifier plus both a pre- and post-modifier (*Pre-M Adj and Adj: Raw # / Pre-modifier: Raw # + Both: Raw #*).

5.3.3.21 Post-M *of*: Raw

The number of common head noun phrases with an *of* phrase post-modifier. Calculated by counting the number of non-blank cells in the Post-M *of* column. Examples include *a state of, the same amount of, fines of, a part of, and the take-over of*.

5.3.3.22 Post-M *of*: %

The percentage of common head noun phrases with an *of* phrase post-modifier. Calculated by dividing the raw number of noun phrases with an *of* phrase post-modifier

pattern by the raw number of total noun phrases with a post-modifier plus both a pre- and post-modifier (*Post-M of: Raw # / Post-modifier: Raw # + Both: Raw #*).

5.3.3.23 Post-M ot. Prep: Raw

The number of common head noun phrases with a prepositional phrase post-modifier (other than *of*). Calculated by counting the number of non-blank cells in the Post-M ot. Prep column. Examples include *food for, people with, their shopping habits over, loss in, and violence and destruction on*.

5.3.3.24 Post-M ot. Prep: %

The percentage of common head noun phrases with a prepositional phrase post-modifier that is not an *of* phrase. Calculated by dividing the raw number of noun phrases with a prepositional post-modifier (other than an *of* phrase) by the raw number of total noun phrases with a post-modifier plus both a pre- and post-modifier (*Post-M ot. Prep: Raw # / Post-modifier: Raw # + Both: Raw #*).

5.3.3.25 Post-M app: Raw

The number of common head noun phrases with an appositive post-modifier. Calculated by counting the number of non-blank cells in the Post-M app column. Examples include *Robert Greenwald's film, Wal-Mart: Xxx High Cost of Low Price; mammography a form of X-ray testing of XXX breasts; ACT's chief executive officer, Richard L. Ferguson; students, elementary and secondary; and one home, Earth*.

5.3.3.26 Post-M ot. Prep: %

The percentage of noun phrases modified after the common head noun with an appositive. Calculated by dividing the raw number of noun phrases with an appositive post-modifier by the raw number of total noun phrases with a post-modifier plus both a pre- and post-modifier (*Post-M app: Raw # / Post-modifier: Raw # + Both: Raw #*).

5.3.3.27 Post-M other: Raw

The number of common head noun phrases with a post-modifier that is neither a prepositional phrase (of any kind), an appositive, or a relative clause. Calculated by counting the number of non-blank cells in the Post-M other column. Examples include *big decisions to, a new foundation if, walls in order to, our responsibility to, and the letter because*.

5.3.3.28 Post-M other: %

The percentage of common head noun phrases with a post-modifier other than a prepositional phrase, an appositive, or a relative clause. Calculated by dividing the raw number of noun phrases with an “other” post-modifier by the raw number of total noun phrases with a post-modifier plus both a pre- and post-modifier ($Post-M\ other: Raw\ \# / Post-modifier: Raw\ \# + Both: Raw\ \#$).

5.3.3.29 Post-M Rel Cl: Raw

The number of common head noun phrases with a relative clause post-modifier. Calculated by counting the number of non-blank cells in the Post-M Rel Cl column. Examples include *more classes which, a problem when, the less time [that], a major problem that, and people who*.

5.3.3.30 Post-M Rel Cl: %

The percentage of common head noun phrases with a relative clause post-modifier. Calculated by dividing the raw number of noun phrases with a relative clause post-modifier by the raw number of total noun phrases with a post-modifier plus both a pre- and post-modifier ($Post-M\ Rel\ Cl: Raw\ \# / Post-modifier: Raw\ \# + Both: Raw\ \#$).

5.3.3.31 Omit relativizer: Raw

The number of common head noun phrases with a relative clause post-modifier in which the relativizer is omitted. Calculated by counting the number of non-blank cells in the

omit relativizer column. Examples include *the way [that]*, *the damage [that]*, *the less time [that]*, *issues [that]*, and *the electricity [that]*. The omitted relative clauses were identified when manually identifying the noun phrases in each essay; upon identification, [that] was added to the phrase for later categorization and classification.

5.3.3.32 Omit relativizer: %

The percentage of common head noun phrases with a relative clause post-modifier in which the relativizer is omitted. Calculated by dividing the raw number of noun phrases with an omitted relativizer by the raw number of total noun phrases with a relative clause post-modifier (*Omit Relativizer: Raw # / Post-modifier Rel Cl: Raw #*).

5.3.3.33 Which relativizer: Raw

The number of noun phrases with a relative clause post-modifier beginning with *which*. Calculated by counting the number of non-blank cells in the *which* relativizer column. Examples include *humanity's biggest problem which*, *the home which*, *coal and petroleum which*, *decisions which*, and *international dimensions which*.

5.3.3.34 Which relativizer: %

The percentage of noun phrases with a relative clause post-modifier beginning with *which*. Calculated by dividing the raw number of noun phrases with a *which* relativizer by the raw number of total noun phrases with a relative clause post-modifier (*Which Relativizer: Raw # / Post-modifier Rel Cl: Raw #*).

5.3.3.35 That relativizer: Raw

The number of noun phrases with a relative clause post-modifier which begins with *that*. Calculated by counting the number of non-blank cells in the *that* relativizer column. Examples include *the story that*, *all evidences that*, *belief that*, *some people that*, and *two different articles that*.

5.3.3.36 *That* relativizer: %

The percentage of noun phrases with a relative clause post-modifier which begins with *that*. Calculated by dividing the raw number of noun phrases with a *that* relativizer by the raw number of total noun phrases with a relative clause post-modifier (that *Relativizer: Raw # / Post-modifier Rel Cl: Raw #*).

5.3.3.37 Data Quality Control

Four columns on the summary worksheet ensured the accuracy of the manual classification. The first quality control column ensured that all of the noun phrases were categorized, with none being inadvertently left out; to this end, a formula was created to display “yes” if the sum of the numbers in the no modifier, pre-modifier, post-modifier, and both columns equalled the total number of noun phrases. Since all of the noun phrases were separated into those four columns, a display of “no” indicated there was an error in categorizing the data (namely a noun phrase had not been categorized or had been categorized more than once), and that worksheet had to be re-examined for errors in analysis. While the “yes/no” formula ensured accuracy of raw numbers, the second quality control column ensured accuracy of the percentages being calculated. The sum of the percentages of no modifier, pre-modifier, post-modifier, and both had to equal 100%. Because, as with the raw number quality control, all of the noun phrases were separated into those four columns, if the percentage check did not display 100%, a noun phrase was either not categorized or had been categorized multiple times, causing the data to be re-examined for errors.

As discussed above, all phrases with post-modifiers, including those with both pre- and post-modifiers, were further categorized into *of*-phrase post-modifier, other prepositional phrase post-modifier, appositive post-modifier, other post-modifier, and relative clause post-modifier columns. Since all noun phrases with post modifiers had to be categorized into one

of these four categories, the sum of the percentages, if no errors were present, had to be 100%. The third quality control column contained a formula which automatically displayed the aforementioned percentage. If the percentage was less than 100%, one or more post-modifiers had not been categorized; if the percentage was more than 100%, one or more post-modifiers had been categorized more than once. Any display other than 100% caused the data in that worksheet to be categorized again. The noun phrases that contained pre-modifiers, including those that had both pre- and post-modifiers, were further categorized into noun pre-modifier, adjective pre-modifier, and adjective + *and* adjective pre-modifier columns. To ensure all phrases with pre-modifiers were further categorized, a formula compared the sum of the pre-modifier breakdown (noun pre-modifier, adjective pre-modifier, and adjective + *and* adjective pre-modifier) to the sum of the pre-modifier and both columns. If the two sums were equal, as they should be if no errors were made, then “yes” was displayed in this last quality control column. If “no” was displayed, a pre-modified noun phrase was either not categorized or had been categorized multiple times, causing the data to be re-examined for errors.

These quality control columns ensured no errors in the quantitative analysis were involved in the classification of the noun phrases. While it was not possible to create formulas which could verify the qualitative analysis involved in categorizing and classifying the noun phrases [e.g. ensuring that the common head noun phrase *two different articles that* was first categorized to the both (pre- and post-modifier) column, further categorized with an adjective pre-modifier and a relative clause post-modifier, then further classified as *that* relativizer], each noun phrase was viewed and read twice: once for the initial categorization to no modifier, pre-modifier, post-modifier, both, or proper noun categories, then viewed and read again from each of those categories to further classify; this double view allowed errors that had been made in the initial categorization to be caught and rectified. For example, if the

noun phrase *courses and curriculum* was wrongly classified into the pre-modifier column, upon further categorization into the noun pre-modifier, adjective pre-modifier, or adjective + *and* adjective pre-modifier category, the error would be noticed, and the noun phrase would be moved to the no modifier category. Similarly, if the noun phrase *all the people* was initially classified into the Both (pre- and post-modifier) column, upon the further categorization into the *of*-phrase, other prepositional phrase, appositive, other, or relative clause category, the error would be noticed and rectified by moving the noun phrase to the pre-modifier column.

5.4 Summary

This study aims to confirm differences between freshman composition and scholarly writing in order to establish a framework which can inform a more appropriate instructional model for novice writers, particularly those studying EAP in the IEP. This chapter detailed the methods employed to complete that aim. In providing an overview of corpus-cited references, I described how the references can be useful to both instructors and researchers. One such example is the scholarly writing profile presented in section 5.2 which instructors can exploit to inform teaching materials and researchers can explore to identify paths of inquiry. Section 5.2 detailed the methods used to approach the corpus-cited references in order to gather data about scholarly writing and explained how these approaches and the creation of the scholarly writing profile led to the identification of the noun phrase as a point of investigation between scholarly writing and freshman composition. Section 5.3 then detailed the methods used in the NAFWiC to identify, classify, and analyze the noun phrase in order to make that comparison. Comparative results are presented in Chapter 6.

CHAPTER SIX

COMPARATIVE RESULTS AND DISCUSSION

6.1 Introduction

Corpus-based research describing actual patterns of language use have steadily increased over the past several decades, especially research describing the register of academic writing. While these studies provide valuable insight into the choices expert writers make, their findings are not appropriate targets for novice writers. Descriptions of expert writing, or even that of graduate writing, may not be valid for first year undergraduate writing or those students seeking to embark on university studies. Chapter 2 section 2.4.1 provides a full discussion on this issue. One of the most important reasons is, as Bazerman (2006, p. 27-28) stresses, if we require novice writers to attempt to take on expert writing “we may put them into positions too distant from their current selves for them to make sense of.” Chapter 2 proposed a staircase model to inform more appropriate instructional materials for novice writers, particularly those at level 1 of the staircase, EAP writers in the IEP. Materials informed by the model are more likely to present writing that can be achievable for students, reflects their communicative purposes, but is also acceptable to the gatekeepers of the discourse, thereby allowing learners “to engage with and grow into disciplinary possibilities.” To confirm differences between scholarly writing and freshman composition writing, and serve as an informant to such materials, this chapter reports comparative results on the use of noun phrases in the two levels of writing. Specifically, section 6.2 discusses differences in the distribution of lexical word classes, section 6.3 explores differences in the modified common head noun phrase, and section 6.4 presents a schematic representation of the noun phrase for both levels of writing. As this study aims to inform EAP writing instruction, pedagogical implications of these findings are discussed in Chapter 7.

It is worth noting that the results included in this chapter are important as they reveal distinct differences between expert and novice writing that have heretofore gone unnoticed. These substantial differences emphasize the argument in this thesis that scholarly writing is a poor model for EAP writers.

6.2 Lexical Word Classes

6.2.1 Distribution of lexical word classes

As discussed in Chapter 5, the most salient linguistic feature of scholarly writing is its extensive use of nouns and noun phrases. Given that noun phrases specify what a text is about, their frequent use is no surprise. Nearly 300,000 (per million) of all content words in scholarly writing are nouns, compared to only ~100,000 (per million) adjectives and verbs and less than ~25,000 (per million) adverbs (Biber, et al, 1999, p. 65). As discussed in Chapter 5, this phenomenon led to the interest in and recognition of the importance of the noun phrase in scholarly writing and its investigation in this thesis. Figure 6.1, reproduced from Chapter 5, displays the distribution of lexical word classes across registers. Figure 6.2 displays the distribution of lexical word classes of successful first year composition writing as represented in the NAFWiC. To facilitate comparison, the data from the two figures have been combined and presented as percentages in Figure 6.3 and normed per million in Figure 6.4.

Figure 6.1 Distribution of Lexical Word Classes across Registers (Biber et al, 1999, p. 65)

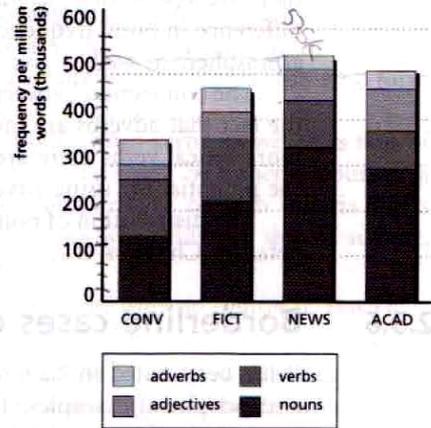


Figure 6.2 Distribution of Lexical Word Classes in the NAFWiC

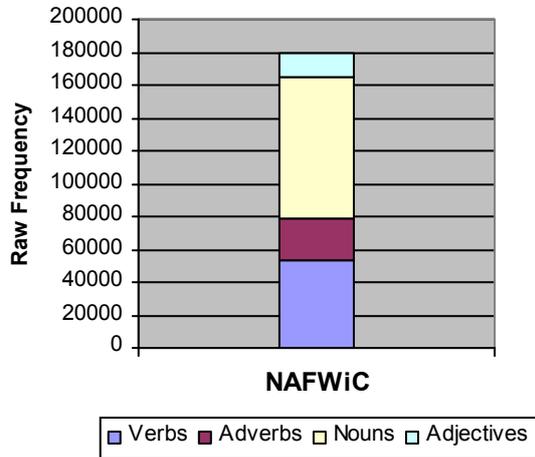


Figure 6.3 Distribution (percentage) of Lexical Word Classes in Scholarly Writing and the NAFWiC

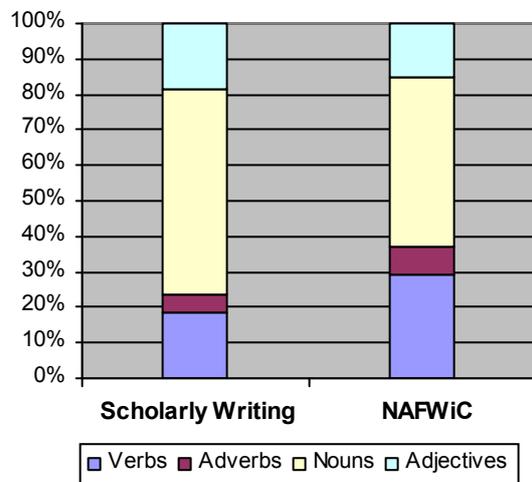
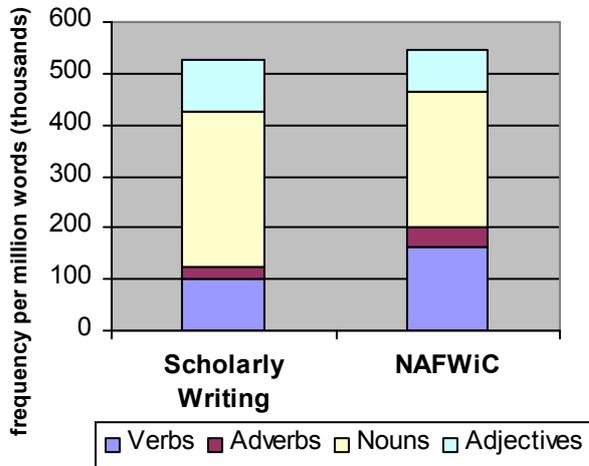


Figure 6.4 Distribution (normed per million) of Lexical Word Classes in Scholarly Writing and the NAFWiC



In the NAFWiC, nearly 50% of content words in scholarly writing are nouns, compared to nearly 30% verbs, 15% adjectives, and 8% adverbs. Compared to scholarly writing, this is 11% fewer instances of noun usage and 10% more instances of verb usage. Not surprisingly, fewer instances of adjectives (nearly 5%) accompany fewer instances of nouns, and more instances of adverbs (nearly 3%) accompany more instances of verbs. Table 6.1 presents a comparison of the distribution of lexical word classes in scholarly writing versus the NAFWiC in percentages.

Table 6.1 Comparison (percentage) of Lexical Word Classes in Scholarly Writing and the NAFWiC

	Scholarly Writing	NAFWiC
Verbs	19	29.49
Adverbs	5	7.79
Nouns	59	48.13
Adjectives	19	14.59

6.2.2 Discussion

Though the use of nouns compared to other lexical word classes in successful freshman composition writing is not quite as striking as that in scholarly writing—in fact, a closer look at Figure 6.1 shows that distribution of lexical word classes in the NAFWiC most closely mirrors that of fiction—nouns are still the most prominent lexical word choice driving

successful first year composition writing. This bears significance for a number of reasons. Firstly, it confirms the investigation of noun phrases in this study is a worthy endeavor. Secondly, it demonstrates that there is some similarity in the two levels of writing in the same register. Thirdly, it hints at the fact that differences exist, and those differences may be important. Following the investigation of the use of noun phrase modifiers and modifier types, differences found between scholarly and first year writing are reported and discussed below.¹

6.3 The Noun Phrase

6.3.1 The definite article

6.3.1.1 Distribution in Scholarly Writing

Two elements are required for a noun phrase: a determiner and a noun. Determiners are function words which are used to specify the reference of a noun (Biber, et al, 1999, p. 258), the most common of which are the definite and indefinite articles. As discussed in Chapter 5, the definite article is highly employed by expert writers. In fact, Figure 6.5 shows that the definite article is used more in scholarly writing than in any other register; it is also the most common definite determiner as shown in Figure 6.6 (reproduced from Chapter 5).

¹ The raw analysis of the noun phrases by subcorpus can be found in Appendix 8 on CD-Rom.

Figure 6.5 Distribution of Articles across Registers (Biber et al, 1999, p. 267)

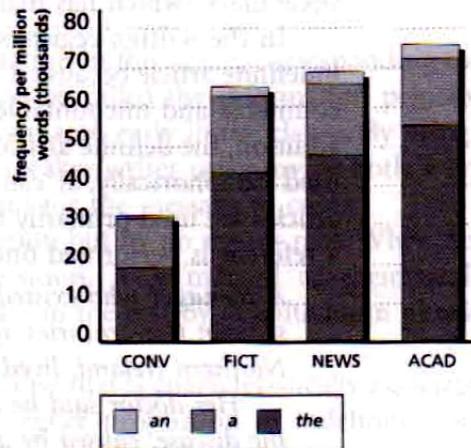
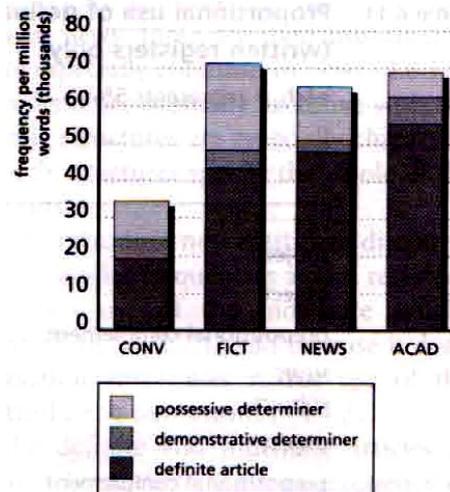


Figure 6.6 Distribution of Definite Determiners across Registers (Biber et al, 1999, p. 270)



In scholarly writing, the definite article is used 85% of the time in the subject position and nearly 80% of the time as a definite determiner (Biber et al, 1999, p. 269-270). Figure 6.7, reproduced from Chapter 5, displays the use of definite and indefinite articles in different syntactic roles across the written registers of the *LGSWE*, highlighting the preference for *the* in the subject position; this is logical, given that in English, the subject is regularly placed early in the clause, and the definite article provides a succinct way of specifying reference. As such, Figure 6.8 shows that cataphoric reference is the most common use of the definite article in scholarly writing (Biber, et al, 1999, p. 267).

across the NAFWiC. Closer investigation, however, reveals three outliers: the persuasive subcorpus (6.46%), summary synthesis subcorpus (9.21%), and profile subcorpus (9.36%). As outliers may be indicative of data that belong to a different population than the rest of the sample set, the presence of these outliers indicates the potential for a distinction between these text types. As is explored below, however, this is not the case overall.

Table 6.2 Frequency of *the* in the NAFWiC organized by subcorpus

<i>Subcorpus</i>	<i>Raw #</i>	<i>%</i>
THEM	337	29.03
REMB	133	28.30
PROF	206	9.36
NARR	383	17.72
LITR	186	32.69
FILM	121	27.38
EXPO	899	28.42
CHAR	364	27.85
ANAL	1060	31.43
PERS	1332	27.48
SUMM	227	9.21
PROP	487	6.46
TOTAL	5735	
AVERAGE		20.35

6.3.1.3 Discussion

Table 6.3 displays a comparison of the use of the definite article in common head noun phrases in scholarly writing and the NAFWiC. Even considering the median² of 27.43%, use of *the* is considerably less in the NAFWiC than in scholarly writing. An average use of 20.34% (or median use of 27.43%) of the definite article among subcorpora in the NAFWiC is not high enough to identify *the* as a salient aspect of noun phrases in successful first year composition writing while it is a firm aspect in scholarly writing.

Table 6.3: Frequency Comparison of *the* in Common Head Noun Phrases in the NAFWiC Versus Scholarly Writing

<i>the</i>	NAFWiC	Scholarly Writing
%	20.35	85
(mean)	(27.43%)	

As mentioned above, the definite article in scholarly writing is often used in cataphoric reference to introduce ideas to the reader; a definite noun phrase can also mark the mention of important information that was introduced earlier in a text (Biber, et al, 1999, p. 579, 269). In scholarly writing, the writer bears the responsibility to inform the reader of connections within the text, and across paragraph boundaries, use of the definite article in a noun phrase is often used to create cohesion (Carter & McCarthy, 2006, p. 265). Additionally, scholarly writing assumes little to no shared context between the writer and the reader. In contrast, forms in conversation reflect a heavy reliance on shared situation (Biber, et al, 1999, p. 579), and, as such, employ fewer definite articles. The differences in purpose of expert and student writing discussed in Chapter 2—writing to share new information versus writing to demonstrate knowledge—could explain the substantial difference in the frequency of *the* in the two levels of writing. The shared context present in student writing—the course and

² The median refers to the middle number in a data set; the median is more representative than the average (or mean) when outliers are present, as is the case here.

assignment—missing from scholarly writing could also account for the difference. While different purposes and audiences of the two levels of writing indicate potential causes for this difference, it may not be possible to identify an exact cause; nevertheless, the difference is certainly noteworthy. As noted in Chapter 5, those structures which are in contrasting use in register can be important foci for instruction; rather than noting the difference among register, however, the differences could be noted at the level of genre and taught as a feature of student writing. This idea is explored further in section 6.4.3.

6.3.2 *Pre-modifiers*

6.3.2.1 Distribution in Scholarly Writing

As noted in Chapter 5, nearly 60% of nouns in scholarly writing are modified. Of those, approximately 25% have a pre-modifier (see Figure 6.9); adjectives and nouns are the main types of noun pre-modification in scholarly writing. Figure 6.10 below displays frequency of pre-modifier types across registers. Around 70% of pre-modifiers in scholarly writing are common adjectives; approximately 30% are nouns. The remaining 10% of pre-modifiers are *ed-* or *ing-*adjectives; although neither of these types of pre-modifiers are very frequent in scholarly writing, they are somewhat more frequent in scholarly writing than in other registers. Pre-modifiers are typically used in scholarly writing with referents; that is, to introduce or refer to an idea within the text (Biber, et al, 1999, p. 585). While pre-modifiers can be used for a first, second, or subsequent mention of a referent, they are used more often for subsequent mentions (40%, compared to 32% first mention and 28% second mention) (Biber, et al, 1999, p. 586).

Figure 6.9 Distribution of Noun Phrases with Pre-Modifiers and Post-Modifiers (Biber et al, 1999, p. 578)

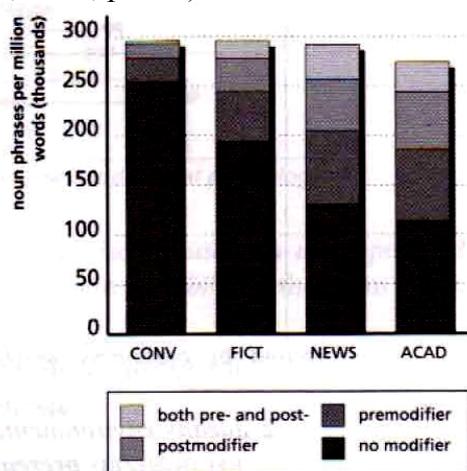
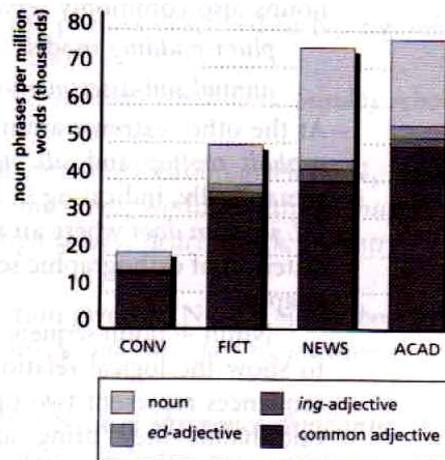


Figure 6.10 Frequency of pre-modifier types across registers (Biber, et al., 1999, p. 589)



As shown in Figure 6.11 below, nearly 80% of pre-modified noun phrases in scholarly writing have a single word pre-modifier, and another nearly 20% have two word pre-modifiers. Only about 2% of noun phrase pre-modifiers in scholarly writing are three or four words in length; of those, nearly all are coordinated pre-modifier type, either the adjective + *or* + adjective (25%) or adjective + *and* + adjective (75%) pattern (Biber, et al., 1999, p. 601) pattern. The adjective + *and* + adjective pre-modifier type typically assumes an evaluative role, offering an opinion or stance on the noun being modified (Carter & McCarthy, 2006, p. 268) or specifically identifies two separate characteristics of a single referent or two mutually exclusive attributes of a plural or uncountable referent (Biber, et al, 1999, p. 601). As shown in Figure 6.12, the use of coordinated pre-modifiers is limited mainly to scholarly writing (Biber, et al., p. 601; Carter & McCarthy, 2006, p. 268). Both the coordinated and the *ed*- or *ing*-adjective pre-modifier types are consistent with the category of those features in the profile presented in Chapter 5 which are not necessarily common in scholarly writing but are typically only found there; structures in this category can be an important focus of instruction because students may not encounter them otherwise.

Figure 6.11 Distribution of pre-modification by length (Biber, et al, 1999, p. 597)

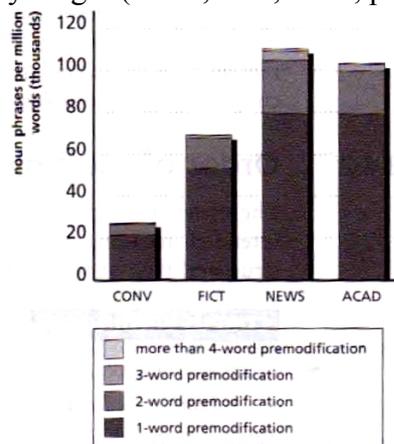
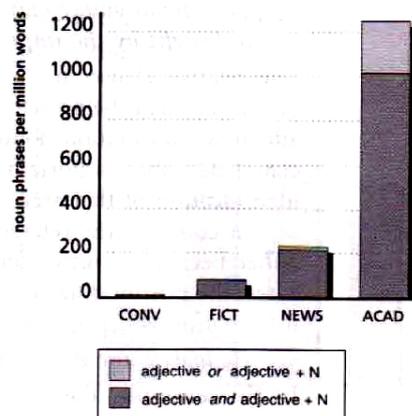


Figure 6.12 Frequency of coordinated pre-modifiers across registers (Biber, et al., p. 601)



6.3.2.2 Distribution in the NAFWiC

In the 246 texts of the NAFWiC, 10,038 of the 28,187 common head noun phrases have a pre-modifier; that is, 35.61% of common head noun phrases in the NAFWiC are pre-modified. Table 6.4 below displays the frequency of common head noun phrases with a pre-modifier according to subcorpus. Table 6.5 displays the frequency of pre-modifier types, also by subcorpus. Overall, pre-modification in the NAFWiC, as well as pre-modification type, is consistent regardless of text type.³

Table 6.4: Frequency of noun phrases with a pre-modifier in the NAFWiC organized by subcorpus

<i>Subcorpus</i>	<i>Raw #</i>	<i>%</i>
THEM	356	30.66
REMB	194	41.28
PROF	279	40.79
NARR	961	44.45
LITR	164	28.82
FILM	167	37.78
EXPO	1174	37.12
CHAR	527	40.32
ANAL	1034	30.66
PERS	1724	35.56

³ Consistency determined by equal mean and median.

SUMM	671	27.22
PROP	2787	37.00
TOTAL	10038	
AVERAGE		35.61

Table 6.5: Frequency of pre-modifier types in the NAFWiC organized by subcorpus

<i>Subcorpus</i>	<i>Pre-M Noun</i>		<i>Pre-M Adj+and+Adj</i>		<i>Pre-M Adj</i>	
	Raw #	%	Raw #	%	Raw #	%
THEM	109	17.72	2	0.33%	504	81.95%
REMB	22	6.94	4	1.26%	291	91.80%
PROF	46	10.24	3	0.67%	400	89.09%
NARR	65	4.81	13	0.96%	1274	94.23%
LITR	51	16.19	4	1.27%	260	82.54%
FILM	46	16.25	2	0.71%	235	83.04%
EXPO	188	9.93	13	0.69%	1692	89.38%
CHAR	81	10.51	2	0.26%	688	89.23%
ANAL	266	15.16	10	0.57%	1479	84.27%
PERS	459	16.55	19	0.69%	2295	82.76%
SUMM	179	15.12	11	0.93%	994	83.95%
PROP	670	15.57	74	1.72%	3559	83%
TOTAL	2182	13.63	157	0.98%	13671	85.39%

6.3.2.3 Discussion

Table 6.6 displays a comparison of the use of pre-modifiers in common head noun phrases in the NAFWiC with that of scholarly writing. As discussed above, approximately 25% of noun phrases in scholarly writing have a pre-modifier. With pre-modification occurring 35.61% of the time in the NAFWiC, the use of pre-modification in first year composition writing is clearly more frequent than in scholarly writing. Because the complexity of noun phrases tends to increase in correlation to formality (i.e. less in conversation most in scholarly writing), it is surprising that pre-modification occurs in the NAFWiC with more frequency than it does in scholarly writing. This is especially so given the structural complexity of modified noun phrases; “new information presented in academic

texts is packaged as modifiers in noun phrases, resulting in a very high density of information” (Biber, et al, 1999, p. 579). This is certainly something we would expect to see more of in expert rather than novice writing.

One of the differences in pre-modification between the NAFWiC and scholarly writing is not as surprising; although the adjective is the most common pre-modifier in both types of writing, the adjective pre-modifier is substantially more common in the NAFWiC. This could indicate that expert writers have and employ a greater range of pre-modification “tools” than novice writers. This possibility could be investigated among the four levels of writing: does the use of pre-modifier types used in academic writing increase with writing proficiency?

Table 6.6: Frequency comparison (percentage) of pre-modifier types in successful first year composition writing versus scholarly writing

<i>Pre-Modifier</i>	<i>NAFWiC</i>	<i>Scholarly Writing</i>
Pre-modifier	35.61	25
Pre-M Noun	13.63	30
Pre-M Adj	85.39	70
Pre-M Adj <i>and</i> Adj	.98	2

Overall, both scholarly writing and freshman composition writing make frequent use of pre-modification, though expert writers employ a greater range of structures to do so.

6.3.3 *Post-modifiers*

6.3.3.1 Distribution in Scholarly Writing

As shown in Figure 6.9 above, approximately 20% of modified nouns in scholarly writing have a post-modifier. Although post-modifiers, like pre-modifiers, can be used for a first, second, or subsequent mention of a referent, post-modifiers in scholarly writing are typically used for a first mention (51%, compared to 24% for second and subsequent mentions) (Biber, et al, 1999, p. 586). Post-modifying structures of common head noun

phrases are mainly phrasal (see Table 6.12); prepositional phrases are, by far, the most common post-modifier in scholarly writing, comprising approximately 75-80% of all post-modifiers (Biber, et al, 1999, p. 606; Carter & McCarthy, 2006, p. 269). Appositives comprise another 5-7% of post-modifiers. The use of appositive post-modifiers in scholarly writing is different from other registers, however, in that 65% of the time appositives in scholarly writing refer to a proper noun or technical name—rarely human reference—to introduce an acronym or provide an explanation to a technical reference (Biber, et al, 1999, p. 640).

Figure 6.13 Frequency of prepositional phrase post-modifiers across registers (Biber, et al., 1999, p. 606)

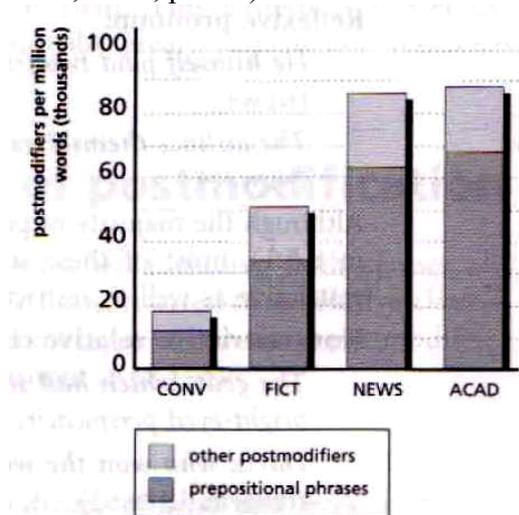


Table 6.7 Prepositions as noun phrase post-modifiers in scholarly writing (Biber, et al., 1999, p. 635)

Preposition	% post-modifier
<i>of</i>	60-65
<i>in</i>	8-10
<i>for</i>	3-5
<i>on</i>	3-5
<i>to</i>	3-5
<i>with</i>	3-5
<i>about</i>	1
<i>at</i>	1
<i>between</i>	1
<i>by</i>	1
<i>from</i>	1
<i>like</i>	1

As illustrated in Table 6.7, six prepositions account for nearly 90% of all those prepositional phrase post-modifiers: *of*, *in*, *for*, *on*, *to*, *with* (Biber, et al., 1999, p. 635). Prepositional phrases are often used in noun phrases to integrate a maximum amount of information; 85% of *of*-phrases present new information (Biber, et al, 1999, p. 305; Carter & McCarthy, 2006, p. 270). Given that this is the general purpose of scholarly writing, it is not surprising to see that *of* accounts for 60-65% of all prepositional phrase post-modifiers; *of*-

phrases are also highly frequent in scholarly writing because of the vast range of functions they complete (Biber, et al, 1999, p. 635). Combined with six additional, though much less frequent prepositions, 12 prepositions account for nearly all prepositional post-modifiers in scholarly writing. Due to the sizeable number of prepositions in English, and their tendency to be confusing for students, a small concentrated list for academic writers could certainly prove useful.

As illustrated in Figure 6.14, however, clauses are also used in post-modification; *to*-clauses are relatively rare in scholarly writing, but *ed*-clauses comprise approximately 3-5% of post-modifiers, *ing*-clauses 10%, and relative clauses another 10-12%. In the approximately 10-12% post-modifying relative clauses in scholarly writing, *which* is by far the most frequent relativizer, used in approximately 47% of relative clause post-modifiers; *that* is the second most frequent relativizer, used in approximately 23% of relative clause post-modifiers (Biber, et al., 1999, p. 611). Omission of the relativizer is least common in scholarly writing, with zero relativizer used in approximately 11% of relative clause post-modifiers (Biber, et al., 1999, p. 609-611). Figure 6.15 displays the frequency of relativizers in post-modifiers in scholarly writing.

Figure 6.14 Frequency of Non-Prepositional Post-modifiers across Registers (Biber, et al., 1999, p. 606)

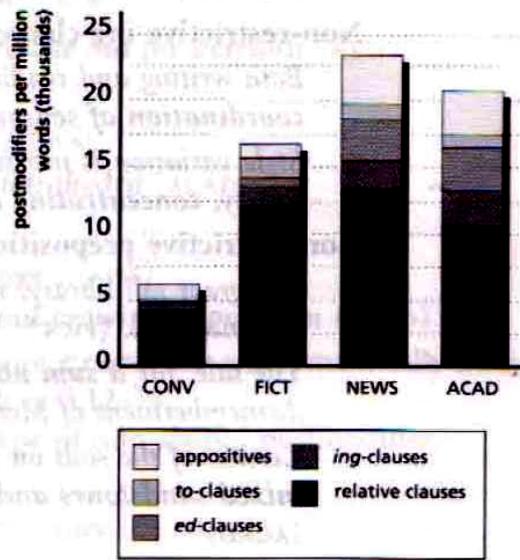
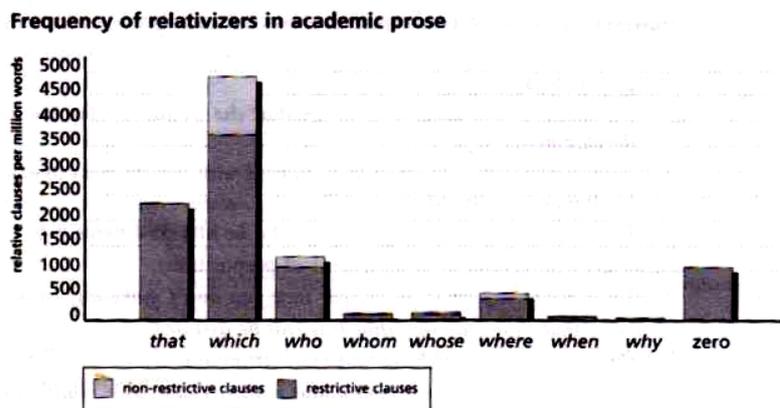


Figure 6.15 Frequency of Relativizers in Scholarly Writing (Biber, et al., 1999, p. 611)



6.3.3.2 Distribution in the NAFWiC

As discussed above, in the 246 texts of the NAFWiC, 5,608 of the 28,187 common head noun phrases have a post-modifier; that is, 19.90% of common head noun phrases in the NAFWiC are post-modified. Table 6.8 below displays the frequency of noun phrases in the NAFWiC that have a post-modifier according to subcorpus.

Table 6.8 Frequency of noun phrases with a post-modifier in the NAFWiC

Subcorpus	Raw #	%
THEM	260	22.39
REMB	60	12.77

PROF	104	15.20
NARR	421	19.47
LITR	150	26.36
FILM	87	37.78
EXPO	711	22.48
CHAR	235	17.98
ANAL	725	21.49
PERS	940	19.39
SUMM	535	21.70%
PROP	1380	18.30%
TOTAL	5608	
AVERAGE		19.90%

Although post-modification occurs in a range of frequencies across subcorpora, there are no outliers,⁴ indicating consistent use of post-modification regardless of text type; similarly the type of post-modification used is consistent across subcorpora. Table 6.9 displays the frequency of post-modifier types according to subcorpus.

Table 6.9: Frequency of post-modifier types in the NAFWiC

<i>Subcorpus</i>	<i>Post-M of</i>		<i>Post-M Other Prep</i>		<i>Post-M App.</i>		<i>Post-M Other</i>		<i>Post-M Relative Cl.</i>	
	Raw #	%	Raw #	%	Raw #	%	Raw #	%	Raw #	%
THEM	143	27.55	194	37.38	9	1.73	84	16.18	89	17.15
REMB	39	21.31	67	36.61	9	4.92	29	15.85	39	21.31
PROF	61	22.26	98	35.77	10	3.65	38	13.87	67	24.45
NARR	156	19.21	284	34.98	9	1.11	148	18.23	215	26.48
LITR	96	31.89	103	34.22	4	1.33	50	16.61	48	15.95
FILM	50	24.63	76	37.44	1	0.49	32	15.76	44	21.67
EXPO	473	33.08	459	32.10	13	0.91	224	15.66	261	18.25
CHAR	137	28.60	145	30.27	13	2.71	80	16.70	104	21.71
ANAL	565	39.07	422	29.18	15	1.04	223	15.42	221	15.28
PERS	514	25.84	730	36.70	16	0.80	354	17.80	375	18.85
SUMM	317	30.25	356	33.97	23	2.19	160	15.27	192	18.32
PROP	857	29.59	1127	38.92	20	0.69	360	12.43	532	18.37
TOTAL	3408	29.43	4061	35.07	142	1.23	1782	15.39	2187	18.89

⁴ Based on similarity of the median and mean.

Relative clauses account for nearly 20% of post-modifier types in the NAFWiC.

Table 6.10 displays the frequency of relativizer according to subcorpus.

Table 6.10 Frequency of relativizer in post-modifier relative clauses in the NAFWiC

<i>Subcorpus</i>	<i>Zero relativizer</i>		<i>which relativizer</i>		<i>that relativizer</i>	
	Raw #	%	Raw #	%	Raw #	%
THEM	16	17.98	3	3.37	47	52.81
REMB	9	23.08	8	20.51	16	41.03
PROF	13	19.40	2	2.99	34	50.75
NARR	43	20.00	5	2.33	114	53.02
LITR	14	29.17	1	2.08	24	50.00
FILM	11	25.00	0	0.00	29	65.91
EXPO	29	11.11	9	3.45	179	68.58
CHAR	32	30.77	1	0.96	55	52.88
ANAL	42	19.00	9	4.07	132	59.73
PERS	47	12.53	10	2.67	219	58.40
SUMM	36	18.75	9	4.69	110	57.29
PROP	78	14.66	41	7.71	302	56.77
TOTAL	370		98		1261	
AVERAGE		16.92		4.48		57.66

Some variation in the choice of relativizer in post-modifiers is evident across text type, specifically in terms of *which* or *that*. There are no uses of *which* as a relativizer in the film analysis corpus and a less than 1% use of *which* relativizer in the character analysis subcorpus; yet, over 20% of relativizers in the remembering an event subcorpus are *which*. The infrequent use of *which* in the character analysis subcorpus may be easily explained; as discussed in Chapter 4, assignments in the character analysis subcorpus required the analysis of a literary character’s personality, with a person—not an idea or object—generally the subject of the analysis. And while the documentary *Wal Mart: The High Cost of Low Prices* was the subject of the film analysis assignment, the analyses were often centered on the director of, or the people in, the film. The more pressing question may be why the essays in the remembering an event subcorpus use *which* substantially more than the other subcorpora.

Closer examination of the data, however, reveals that this anomaly is due to variation in individual writing style; two files in the remembering an event subcorpus account for all but one use of *which* relativizer in the entire subcorpus. Generally, there is consistency in choice of relativizer across text type in the NAFWiC: a clear preference for *that* relativizer overall, then zero relativizer, and lastly, *which* relativizer.

6.3.3.3 Discussion

Post-modification in scholarly writing and the NAFWiC is similar in several aspects, but overall striking distinctions are also evident between the two levels of writing. The frequency of overall post-modification in scholarly writing and the NAFWiC is virtually equal (20% and 19.90%, respectively), and prepositional phrases are the most common type of post-modifier in both levels of writing (64.5% and approximately 70%, respectively). As illustrated in Table 6.11, however, the use of post-modifier types in the NAFWiC is distinct from that in scholarly writing. Firstly, scholarly writing relies heavily on *of*-phrase prepositions as post-modifiers, but the NAFWiC employs a much greater range of prepositions; only 30% of post-modifiers in the NAFWiC are *of*-phrases (compared to nearly 50% in scholarly writing). As reported in section 6.3.3.1, *of*-phrases are highly productive in scholarly writing post-modification due to the vast number of functions they perform; a lesser reliance on *of*-phrases in the NAFWiC may indicate that novice writers are not yet aware of or able to use all the functions which can be realized by the *of*-phrase post-modifier. However, as also reported in section 6.3.3.1, *of*-phrases are mostly used in scholarly writing to present new information; as discussed in section 6.3.1.3, this purpose of expert writing is not the same as that of student writing, which mainly serves to demonstrate knowledge. This difference in purpose could account for the lesser reliance on *of*-phrases. A study investigating the

function of the *of*-phrase post-modifier in first year undergraduate writing could provide more insight into this difference.

Table 6.11 Frequency comparison (percentage) of post-modifiers and types in scholarly writing and the NAFWiC

<i>Post-Modifier</i>	<i>NAFWiC</i>	<i>Scholarly Writing</i>
Post-modifier	19.90	20
Post-M <i>of</i>	29.43	46
Post-M Other Prep	35.07	24
Post-M App	1.23	7
Post-Mod Other	15.39	8
Post-Mod Rel Cl	18.89	11
Naught Relativizer	16.92	10
<i>which</i> relativizer	4.48	40
<i>that</i> relativizer	52.67	27

As Table 6.11 illustrates, in addition to making use of a wider range of prepositional post-modifiers, the NAFWiC also makes greater use of relative clause post-modifiers (18.89% compared to approximately 11%) and a greater use of other post-modifiers (such as *to* phrase) (15.39% compared to approximately 8%) than scholarly writing. Perhaps the most marked difference between the two levels of writing lies in the choice of relativizer in relative clause post-modifiers. Scholarly writing exhibits a clear preference for *which* relativizer, while the NAFWiC exhibits a clear preference for *that* relativizer, rarely employing *which*. Furthermore, the relativizer is omitted nearly twice as much in the NAFWiC compared to scholarly writing (almost 17% compared to approximately 10%, respectively). Approximately half of relative clauses in conversation which permit omission of the relativizer do so (Biber, et al, 1999, p. 620-621); given the tendency of novice writers to employ casual spoken language to their writing, discussed in Chapter 5 section 5.2.2.2, the frequent omission of the relativizer in the NAFWiC is not particularly surprising.

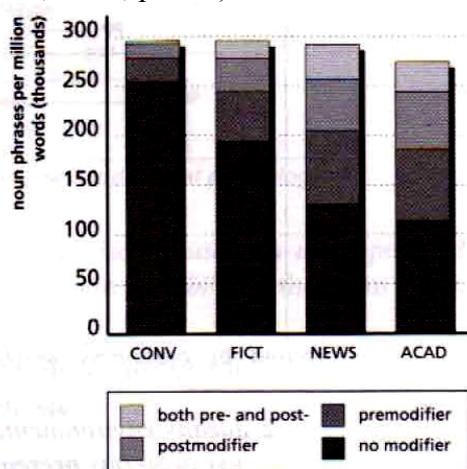
6.3.4 *Non and Dual⁵ Modification*

6.3.4.1 Scholarly Writing

Because of the frequent need for definition and speculation in scholarly writing, it should be no surprise that modified noun phrases are extremely frequent in academic English (Carter & McCarthy, 2006, p. 269). As shown in Figure 6.16, the use of modified noun phrases occurs more often in scholarly writing than any other register (though modified nouns are only slightly less common in News). Only approximately 40% of nouns in scholarly writing do not have a pre- or post-modifier. Not only are nouns regularly modified by a pre- or post-modifier, Figure 6.16 illustrates that a number of noun phrases in scholarly writing have both a pre- and post-modifier (approximately 12%). Like post-modified nouns, dual modified nouns can be used in a first, second, or subsequent mention of a referent, but an overwhelming number of noun phrases that have both a pre- and post-modifier are used for first mention of a referent (70%, compared to 12% for second and subsequent mentions) (Biber, et al, 1999, p. 586); this shows that dual modified nouns have a more restricted role than those with only a pre-modifier or post-modifier.

⁵ The term “dual modification” is used here as a more precise way to refer to both pre- and post-modification of a noun phrase.

Figure 6.16 Distribution of Noun Phrases with Pre-Modifiers and Post-Modifiers (Biber et al, 1999, p. 578)



6.3.4.2 The NAFWiC

In the 246 texts of the NAFWiC, 6,569 of the 28,187 common head noun phrases have no modifier; that is 23.31% of noun phrases are unmodified. However, 5,972 of the 28,187 common head noun phrases have both a pre- and post-modifier; that is 21.19% of noun phrases have both a pre- and post-modifier. Table 6.12 below displays the frequency of noun phrases that have either no or dual modification in the NAFWiC. As with the results reported above, there is continuity of use among the subcorpora.

Table 6.12: Frequency of noun phrases with no or dual modification in the NAFWiC according to subcorpus

<i>Subcorpus</i>	<i>No Modifier</i>		<i>Dual Modifier</i>	
	<i>Raw #</i>	<i>%</i>	<i>Raw #</i>	<i>%</i>
THEM	286	24.63	259	22.31
REMB	93	19.79	123	26.17
PROF	131	19.15	170	24.85
NARR	389	17.99	391	18.09
LITR	104	18.28	151	26.36
FILM	72	16.29	116	26.24
EXPO	559	17.67	719	22.73
CHAR	301	23.03	244	18.67
ANAL	893	26.47	721	21.38
PERS	1135	23.41	1049	21.64
SUMM	746	30.26	513	20.81
PROP	1860	24.66	1516	20.10
TOTAL	6569		5972	
AVERAGE		23.31		21.19

6.3.4.3 Discussion

Table 6.13 Frequency comparison (percentage) of non and dual modification in scholarly writing and the NAFWiC

<i>Modification</i>	<i>NAFWiC</i>	<i>Scholarly Writing</i>
No modifier	23.32	40
Dual Modifier	21.19	12

Table 6.13 compares the frequency of nouns with no modifiers as well as nouns with both a pre- and post-modifier in scholarly writing and the NAFWiC. Although the majority of noun phrases in scholarly writing have at least one modifier, only a few have dual modifiers; there are more than three times as many un-modified noun phrases as dual modified noun phrases in scholarly writing. This simply means that while a considerable number of noun phrases use either a pre-modifier or a post-modifier (60%), few use both at the same time (12%). In these terms, the same can be said to be true of the NAFWiC; while an overwhelming number of noun phrases use either a pre-modifier or a post-modifier (80%),

few use both at the same time (20%). The realization of those numbers in parenthesis, however, is important. Scholarly writing has nearly twice as many non-modified noun phrases but fewer than half dual-modified noun phrases than the NAFWiC. Given the different purposes of scholarly writing and student writing combined with use of modification in (1) complex structures (2) to package information and (3) establish context, it is certainly surprising to see not only more modifiers in the NAFWiC, but more complex modifiers.

Why do novice writers (in the NAFWiC) modify noun phrases more than expert writers (in scholarly writing)? One possible explanation, as discussed in the review of research on learner writing in Chapter 3 section 3.5.5, is that L2 writers over-use certain structures; if learner writers over-use modification patterns, this could explain why novice writers, only one step above learner writers on the staircase model, use more modifiers. Yet, Hinkel (2011, p. 529) actually reports that L2 writers “employ fewer lexical and syntactic modifiers of nouns” and “incorporate fewer modifying and descriptive prepositional phrases,” making it less likely that this is a case of over-use by novice writers. Another potential explanation for the greater use of modification in the NAFWiC could be that novice writers employ pre- and post-modifiers in different ways than expert writers. While pre- and post-modification in scholarly writing involves information packaging and introduction and mention of referents, modification in novice writing may perform other functions. A study investigating the function of pre- and post-modifiers in first year undergraduate writing could provide more insight into their frequent use in this level of writing.

6.3.5 Summary

Table 6.14 presents a compilation of Tables 6.3, 6.6, 6.12, and 6.13 as an overall comparison of modified common head noun phrases in scholarly writing and the NAFWiC. To provide a better understanding of the differences observed, this section reviews those

findings discussed in sections 6.3.1.3-6.3.4.3 comparing noun phrases in the two levels of writing. One such observation worth noting at the outset of this summary is that the subcorpora in the NAFWiC, though displaying a range of frequencies, employed each of the structures in the noun phrase with continuity; even though the text types represent different assignments and purposes, overall the data support the idea that similarities among the text types contribute to a genre of student writing consistently used in the freshman composition discipline.

Table 6.14 Frequency comparisons (percentage) of modified noun phrases in scholarly writing and the NAFWiC

Noun Phrase Feature		Scholarly Writing	NAFWiC
Modifier	No modifier	40	23.32
	Pre-modifier Only	25	35.61
	Post-modifier Only	20	19.90
	Dual Modifier	12	21.19
Pre-modifier	<i>the</i>	85	20.35
	Noun Pre-modifier	30	13.63
	Adj Pre-modifier	70	85.39
	Adj <i>and</i> Adj Pre-Modifier	2	.98
Post-modifier	<i>of</i> Post-modifier	46	29.43
	Other prep. Post-modifier	24	35.07
	Relative Cl. Post-modifier	10-12	18.89
	Zero relativizer	10	16.92
	<i>Which</i> relativizer	40	4.48
	<i>That</i> relativizer	27	57.56
	Appositive Post-modifier	7	1.23
	Other Post-Modifier	8	15.39

One notable difference between the two levels of writing is a greater overall use of modification; fewer than 60% of common head noun phrases in scholarly writing are modified compared to more than 75% in the NAFWiC. Another difference, as discussed above, is that noun phrases in scholarly writing have a dual modifier only 12% of the time, but nouns in the NAFWiC have nearly twice that amount (21%). Furthermore, not only are pre-modifiers more frequent in the NAFWiC than in scholarly writing, pre-modifier types are

used differently in the two levels of writing. Another noteworthy difference in pre-modifier types between the two levels of writing involves the use of adjectives as pre-modifiers. Adjectives are used approximately 70% of the time as pre-modifiers in scholarly writing but 85% of the time in the NAFWiC, comprising an overwhelming majority of pre-modifier types there. Additionally, the noun and adjective + *and* + adjective pre-modifier types are used less than half as frequently in the NAFWiC compared to scholarly writing. Although the two levels of writing have an equal percentage of post-modified noun phrases, they make considerably different choices about how to modify them. The NAFWiC uses nearly twice as many relative clauses and other post-modifier types as well as nearly twice as many *that* and zero relativizers than scholarly writing; however, it uses only one quarter the number of *which* relativizers and only a fragment of appositive post-modifiers compared to scholarly writing. Perhaps the most striking difference between the two levels of writing involves the definite article. While *the* is used approximately 85% of the time as the article of choice in modified common head noun phrases in scholarly writing, *the* is used only an average of 20.35% (or median of 27.43%) of the time in the NAFWiC.

6.4 A Schematic Representation of the Noun Phrase

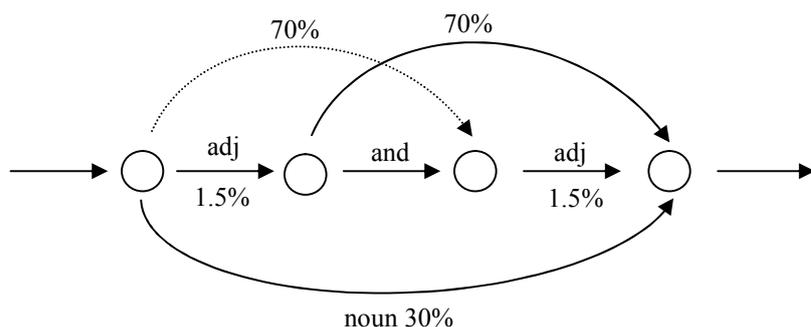
6.4.1 Introduction

A transition network grammar (TNG) is used mainly in computational linguistics as a set of permissible transitions among transformations (e.g. word order, congruence) that can be recognized and analyzed in a given text (Bussman, Kazzazi, & Trauth, 1998, p. 42-43).

TNGs display probabilities of distribution of grammatical choices that computers can automate, for example, in order to apply annotations to a corpus. Figure 6.17 displays a TNG for pre-modification in the common head noun phrase in scholarly writing. The straight arrows represent the path to a pre-modified noun; the circles indicate where a choice will be

made; the arrows also indicate possibilities for those choices (these are the pre-modifier types); the percentages represent the probability of that particular type being chosen. For the TNG below, the first “choice” (circle) is three-fold (indicating three modifier types); the bottom arrow indicates a 30% probability that a noun will be the pre-modifier. Taking that path leads directly to the noun which is to be modified. The dotted arrow indicates a 70% probability that a common adjective will be the pre-modifier. Taking that arrow leads to an adjective and then directly to the noun which is to be modified. The straight arrow indicates a 1.5% probability that *and* will follow the adjective followed by another adjective before going directly to the noun to be modified (adjective + *and* + adjective). TNGs can be constructed based on findings such as those presented in this chapter.

Figure 6.17 Pre-Modifier TNG for Scholarly Writing



While I am not a computational linguist, I see considerable potential value in TNGs offering visual representation of grammatical choices. As mentioned in Chapter 3, corpus-based studies contribute to descriptions of language in use; as Carter and McCarthy (2006, p. 10) note, these descriptions are particularly important for learners to observe and understand how and why speakers use the language they do—not as a prescription for learner use—but as a presentation of data to enable learners to make their own informed choices. Being able to “see” possible choices and the frequency with which those choices are realized can help students and instructors alike develop language skills. In order to display grammatical

choices, and their probabilities, in the noun phrase, I have created a schematic representation (inspired by a TNG) using the findings in Table 6.14 above for both the NAFWiC and scholarly writing. Not only does the schematic representation provide a “clear picture” of the noun phrase in use, it facilitates comparison of use between the two levels of writing.

6.4.2 Schematic Representation for the NAFWIC and Scholarly Writing

Figures 6.17 and 6.18 display the schematic representation of the noun phrase in scholarly writing and the NAFWiC (based on the findings presented in this chapter and compiled in Table 6.14). As discussed above, the figures in the representation display probabilities of distribution of grammatical choices and provide a picture of the noun phrase in each level of writing. As an example, Figure 6.18 shows that 11% of noun phrase post-modifiers are relative clauses, while Figure 6.19 shows that 19% of noun phrase post-modifiers are relative clauses.

Figure 6.18 Schematic Representation of the modified common head noun phrase in scholarly writing

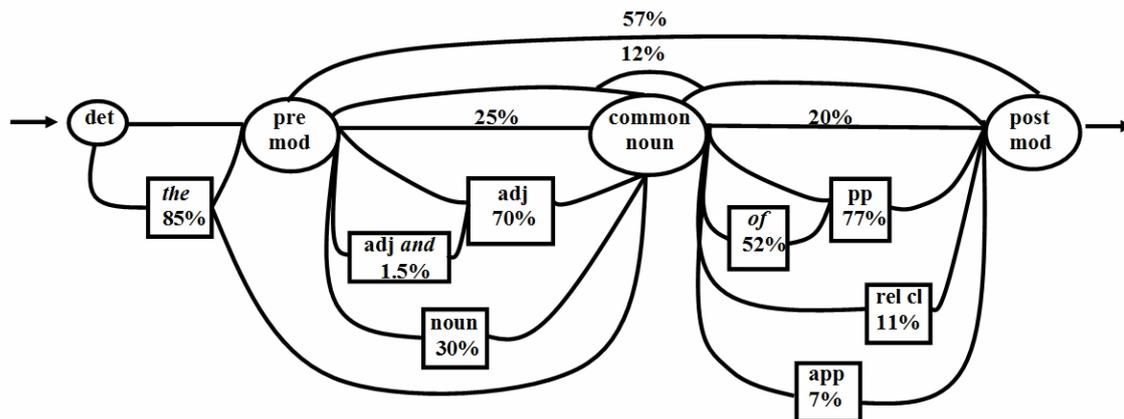
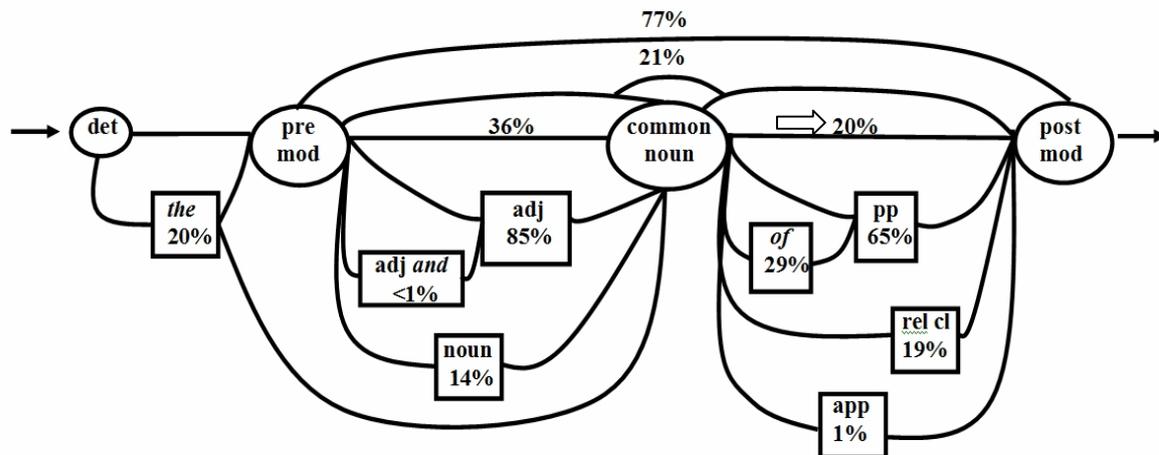


Figure 6.19 Schematic Representation of the modified common head noun phrase in the NAFWIC



6.4.3 Discussion

Clearly noun phrases in both levels of writing employ a discernible pattern. Though there are multiple options in the noun phrase—use of one, the other, or both modifiers, and modifier types—the schematic representations succinctly depict to what extent each choice is made. The figures also clearly depict to what extent the writers at each level make different choices. The only similar frequency in the two representations is highlighted with an arrow (\Rightarrow) on Figure 6.19. As Tribble (2002, p. 138) notes, “although percentages will not give information on statistical significance they provide a useful rough means of differentiating between the texts in question.” These percentages plainly confirm a difference in the NAFWiC and scholarly writing.

In sum, the schematic representation in Figure 6.19 highlights several patterns which characterize the modified common head noun phrase in successful first year composition writing. The NAFWiC

- shows a strong preference for modifying noun phrases (77%);
- makes productive use of prepositional phrases (65%) and relative clauses (19%) as post modifiers (~36%);
- makes prolific use of adjectives as pre-modifiers (85%);
- frequently omits relativizers (17%); and
- prefers the use of *that* as a relativizer (58%).

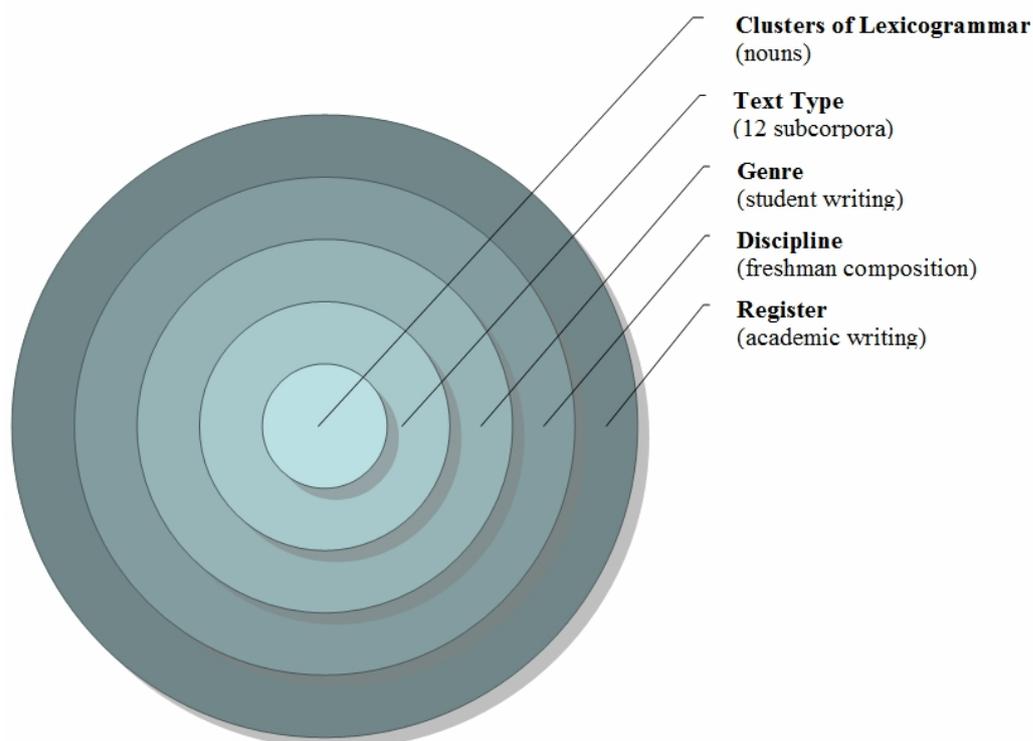
Furthermore, in comparing Figures 6.17 and 6.18, general observations that emerge from the findings which distinguish the NAFWiC from scholarly writing include

- overall greater use of modifiers,
- less reliance on the definite article,
- greater use of common adjective pre-modifiers,

- more diverse use of pre- and post-modifiers types.

Though the two levels of writing do obviously differ in their use of the noun phrase, it should be noted that the differences are likely at the level of genre or discipline. This is explained by considering the relationships among written texts. Figure 6.20, reproduced from Chapter 4, displays text relationships in the NAFWiC. The outer ring signifies the register of academic writing. As a whole, this register is defined by expert writing. In fact, the extremely high use of nouns and modified noun phrases (depicted in the inner ring) is a general characteristic of this register. But as one moves closer to the center of figure, how those characteristics are realized may differ. We already know of differences which exist among the second ring, disciplines, and even the third ring, genres (see Chapter 3). Current corpus-based studies are continuing to explore these differences. The findings in this study, as remarked in section 6.3.5 above, confirm that the various text types used in freshman composition realize grammatical structures similarly; this supports the idea of a genre of student writing. Whether the differences noted here between expert and novice writing fit within genre or discipline, however, is hard to say. It is possible that these differences characterize a distinct freshman composition discipline (which I have argued for in this thesis). But it could also be that these differences characterize student writing as a whole. It could be, too, that both possibilities are true. One way to make such a distinction could be to investigate more of the structures from the profile of scholarly writing (presented in Chapter 5) to see what other differences emerge. Another possibility is to focus on the potentially different functions of the same structure (as was mentioned in sections 6.3.3.3 and 6.3.4.3) to see what differences may emerge.

Figure 6.20 Text Relationships in the NAFWiC



This distinction (difference between genre or discipline) is not as important, however, as that between expert and novice writing. As mentioned above, the findings presented here plainly confirm a difference in the NAFWiC and scholarly writing, thereby confirming the argument that expert writing is a poor model for novice writers.

6.5 Summary

To confirm differences between scholarly writing and first year undergraduate writing, and serve as an informant for instructional materials of the latter, this chapter presented comparative results on the noun phrase in the two levels of writing, specifically in terms of the distribution of lexical word classes and use of noun phrase modifiers and modifier types. The results reported here do confirm that expert and novice writers use diverse structures in differing frequencies in texts, demonstrating that although successful first year composition writing as represented in the NAFWiC shows some resemblance to scholarly writing, overall the two levels of writing use the structures in the modification of common head noun phrases

distinctly. The results and discussion here are important because they reveal differences between expert and novice writing that have heretofore gone unnoticed. A schematic representation displaying the choices that can be made in modifying noun phrases, and the extent to which these choices are typically made at each level of writing, highlights the differences between the NAFWiC and scholarly writing. These differences confirm the arguments that scholarly writing is a poor instructional model for freshman composition, and therefore, freshman composition merits its own description. Chapter 7 explores possible pedagogical applications of the findings presented and discussed here.

CHAPTER SEVEN

PEDAGOGICAL APPLICATIONS

7.1 Introduction

As discussed in Chapter 2, models are not without fault; they do, however, allow potentially excluded participants an opportunity to become involved in academic discourse by facilitating participants' realization of the structures and functions exhibited in the targeted discourse and allowing for eventual independent construction of such discourse through experience and participation. I have argued here that expert writing is not an appropriate model for teaching novice writers: purposes of expert and student writing differ; the expert standard is unfair and descriptively inadequate; the model requires far transfer and thus is ineffective. Yet, a student model may also be problematic: multiple scholars criticize student writing as inauthentic and models of student writing may disregard gatekeepers. This thesis proposes the teaching of English for academic purposes (EAP) writing should be informed by a staircase model of writing progression, thereby addressing the aforementioned concerns. For EAP writers in intensive English programs (IEPs), this target is first year undergraduate writing; more specifically, freshman composition. The schematic representation presented in Chapter 6 of the noun phrase in the NAFWiC highlights some of the features which are characteristic of freshman composition writing; it is these features which should be emphasized for learners seeking to begin undergraduate studies in North American contexts. This chapter explores applications for pedagogical materials given these considerations. Firstly, section 7.2 considers pedagogical materials and their use in classrooms. Section 7.3 follows with a discussion of issues surrounding the integration of findings from corpus-based studies into such pedagogical materials. Finally, section 7.4 surveys current textbooks used in

EAP writing classes in the IEP and outlines applications and recommendations for pedagogical materials based on the proposals and findings presented in this thesis.

7.2 Pedagogical Materials

7.2.1 Definition

In this thesis I have proposed a staircase model to inform more appropriate pedagogical materials for novice writers, particularly those at level 1 of the staircase, EAP writers in the IEP, with the claim that materials informed by the model and use descriptions of student genres are more likely to present writing that can be achievable for students and reflect their communicative purposes, while also being accepted to the gatekeepers of the discourse, thereby allowing learners “to engage with and grow into disciplinary possibilities.” But what, exactly, are “pedagogical materials”? Typically, pedagogical materials are either those created by teachers specifically for his or her own classroom use or published textbooks that are used in a wide-variety of classrooms. While teacher-created materials can be very effective given their targeted relevance to the specific students in a given class, published textbooks are by far the most common type of materials. Applications discussed in section 7.4 are relevant for either teacher-created materials or published textbooks, but while some teachers do create their own materials, nearly every teacher uses published textbooks. Therefore, this discussion centers on “pedagogical materials” as published textbooks.

7.2.2 Development

Textbooks are generally born from one or more motivations; Swales (1995, p. 8) lists six potential motivations for the development and publication of writing textbooks:

1. to pluralize successful and interesting writing activities developed in singular classrooms;

2. to disseminate a vision (or theory or ideology) about the nature of the writing process and how it can be fostered;
3. to demonstrate how a body of knowledge about language or discourse or society can be put to applied use;
4. to combine linguistic, rhetorical and methodological investigations in a genre that easily encompasses all three;
5. to gain prestige and visibility within the chosen discourse community, particularly as an 'all-rounder';
6. to respond to a perceived gap in the market, either as an individual initiative or as a commissioned work.

While any of the above are worthy aims for a textbook, prevailing wisdom holds that “there is no such thing as the perfect textbook” (J. D. Brown, 1995, p. 166). Although textbooks have both benefits and drawbacks in language teaching (for a discussion on these see e.g. J. D. Brown, 1995; Tomlinson, 1998; Graves, 2000; H. D. Brown, 2001; Ferris & Hedgecock, 2005), they hold a firm place in language teaching classrooms for two reasons: many teachers do not have the time or resources to create their own materials for the variety of classes they find themselves teaching and thus need textbooks for instructional materials; many teachers and students feel more confident that a published textbook presents all the important areas which need to be covered in a course (and they would not be able to do so by creating their own materials). Stemming from this last reason, Hyland (1994, p. 252) says, “there is a common assumption that any item in a textbook must be an important learning item and, conversely, that anything not included can be safely omitted from a course.” Therefore, textbooks often play a major role in the development of teachers’ methods and beliefs. What teachers may not understand or take into account, however, is that materials writers make a

range of choices, and these choices often involve compromise (Bell & Gower, 1998). This compromise often results from the need to generalize aspects of the materials so they may be appropriate in a broad range of settings; one particular compromise may often involve presenting statements as accredited facts (Hyland, 1994), which, as discussed below, may falsely portrays language as absolute.

Materials writers also have to make judgments about what language and structures to include in their textbooks (Biber & Reppen, 2002) and should combine institutional practices and knowledge of research (Swales & Feak, 2006) when doing so. Unfortunately, however, many of these judgments, which are often based on the author's beliefs about language and the language skill covered in the materials, are often unconscious, therefore not specifically accounting for research. For example, an author must consider in developing a textbook (or an instructor must consider in adopting one) his or her definition of the skill being taught; in other words, how an author or teacher defines writing or grammar (and the role they should play together) affects their approach to it. Authors must also consider issues of frequency (Biber & Reppen, 2002); but because judgments are often made unconsciously, frequency decisions are often based on intuition, which is often incorrect, and therefore, textbooks "often fail to provide an accurate reflection of the language actually used by speakers and writers in natural situations" (Biber & Reppen, 2002, p. 200) having "serious effects on students' successful acquisition of essential communicative discourse features" (Hyland, 1994, p. 252). In response to this issue, Hyland (1994, p. 252) purports more focused reliance on corpus studies research; similarly, Yoo (2009, p. 276) argues that materials can facilitate the learning process for students by simply consulting existent corpus studies, doing a great disservice if not done so. But this advice is not new; at least not in European contexts, as Dudley-Evans (1988, p. 28) more than two decades ago claimed that "materials writers need

more detailed analyses of rhetorical and linguistic organization of the tasks if they are not to be over-reliant on their own intuition.” With the rise of corpus-based studies providing descriptions of language, textbooks have a growing base of data which can be used to inform their decisions. However, as discussed in section 7.3 below, the findings of corpus-based studies have not had much success in finding their way to textbooks.

7.3 Corpus Findings in Pedagogical Materials

7.3.1 Prospects

Hunston (2002, p. 1) asserts that “it is no exaggeration to say that corpora, and the study of corpora, have revolutionised the study of language, and of the applications of language.” According to O’Keeffe, McCarthy, and Carter (2007, p. 21), “the contribution of corpus linguistics...is difficult to dispute.” And Mukherjee and Rohrback (2006, p. 205) claim “there is no doubt that corpus-linguistic research has exerted an enormous influence on the teaching of English as a Foreign Language (EFL) around the world.” Chapter 3 discussed four areas in which corpus linguistics has made a significant contribution to understanding language use—phraseology, frequency, register description, and applications in language teaching—and specifically proposed three types of publications which have lead to applications in language teaching: those that provide an overview of major language teaching applications, those that provide an overview of the field as a whole, and those that provide practical explanations for realizing those applications in the classroom.¹ The largest and perhaps most successful contribution of corpus linguistics has been in lexicography; non corpus-based learner dictionaries are now “virtually unheard-of” (Hunston, 2002, p. 96).

¹ Although teacher-created materials are acknowledged as a legitimate type of “pedagogical materials,” the discussion in this chapter focuses only on textbooks, as explained in section 7.2.1. This distinction is particularly relevant in this discussion in section 7.3.1 as many teacher-created materials and articles advocating such exist in the literature. (It might should be noted, though, that these materials are more European based, as data-driven learning or other uses of corpora in teacher-created materials are only just being considered in North America.)

Conrad (2000, p. 549) hypothesized that corpus linguistics had “the potential to revolutionize the teaching of grammar.” She predicted that 1) monolithic descriptions of English grammar could be replaced by register-specific descriptions; 2) the teaching of grammar could become more integrated with the teaching of vocabulary; and 3) emphasis could shift from structural accuracy to the appropriate conditions of use for alternative grammatical constructions. The first two hypotheses have certainly been realized in the research, as discussed in Chapter 3; the Academic Word List (AWL) (Coxhead, 2000) and the idea of collocations have also become recognized as important concepts in many textbooks. But few textbooks in the North American context have, over the past decade, even come close to realizing these hypotheses. Notable exceptions include some books based on the MICASE [e.g. *Academic Listening Strategies* (Salehzadeh, 2005), *Academic Interactions* (Feak, Reinhart, & Rohlck, 2009), and *Four Point Listening and Speaking* (Parrish, 2009)]: and the AWL [e.g. *Focus on Vocabulary*, (Schmitt & Schmitt, 2005), *Vocabulary Mastery Series* (Wells & Valcourt, 2010)]. Two additional exceptions are The *Touchstone* series (McCarthy, McCarten, & Sandiford, 2005) (a multi-level integrated skills course book which is completely corpus-based) and *Real Grammar* (Conrad & Biber, 2009) (a supplemental grammar textbook driven by findings in the *LGSWE*). Overall, however, pedagogical materials have not been deeply impacted by corpus findings.

7.3.2 Problems

In his plenary address at the 2005 TESOL convention, Biber stated that two problems plague the progression of corpus linguistics in language teaching; one of them was, not surprisingly, adapting corpus findings to classroom materials. In her statement that corpus linguistics could revolutionize grammar teaching, Conrad (2000) acknowledged that three factors must be considered for these revolutions to occur. Firstly, she recommends that

frequency data alone not dictate pedagogy; frequency does not equal relevance. Secondly, she recommends that materials writers use corpus findings subtly, like asking “students to use the most common verb + *that* clause combinations for the same purposes that corpus research has revealed among native speakers” (Conrad, 2000, p. 557). Finally, she acknowledges that teachers must be willing to depart from a reliance on intuition and traditional grammar syllabi; these last two ideas challenge long-held teaching practices and are certainly, in my experience, the most difficult to overcome². Mukherjee and Rohrback (2006) also point out that there is a widening gap and a growing lag between on-going and intensive corpus-linguistic research on the one hand and classroom teaching on the other. They attribute this lack of collaboration on differing viewpoints, also highlighting a distinction between frequency and relevance. Specifically considering issues with using corpora to inform materials for writing instruction, Gilquin, Granger, and Paquot (2007) discuss relevance, as well; the discipline specific nature of writing and the individual differences among learners (e.g. language level, goals, target language) seem nearly impossible to reconcile. One conclusion these issues highlight is that there is still room for growth for corpus linguistics and corpus-based studies in language teaching.

7.3.3 Potentials

The model proposed here addresses Gilquin et al’s concerns from above; the students in the IEP are all high-intermediate to advanced learners of English who desire to enter a North American institution of higher education and will have to immediately engage in freshman composition—thus, the discipline specific writing, language level, and goals of the

² Several years ago I worked as a corpus researcher for a very large publishing company in North America which publishes the, arguably, most popular grammar series ever. The author of the series was to email me questions that came up during the revision process; for example, which is more frequent: *having* + past participle or *to have* + past participle? and I would use the publisher’s corpus network to investigate. Increasingly, results from the inquires conflicted with the author’s intuition about structures; in response to one conclusion that a certain structure was infrequent, she commented “but I know I hear that!” Ultimately, the author found she had no use for the information in the corpus, and my “services” were only required for approximately half of the revision. (Interestingly enough, the series now claims to be corpus-based.)

students are reconciled. Issues regarding the aspect of frequency and relevancy can also be reconciled. Stubbs (2010) acknowledges that corpus linguists often get caught up in the excitement of empirical findings without considering relevance, but the relevance in that frequency, he proposes, is its characterization of language; for example, those structures which are frequent in a text type are those which characterize it (as argued in Chapter 6). In their recommendations for including corpus-findings in teacher materials, Biber and Reppen (2002, p. 199) address the issues of frequency and relevance; they propose frequency be considered in materials writing in three areas: what structures should be included and to what extent, what order structures should be presented, and what words should be included in those structures. Furthermore, they argue that given the importance of frequency in acquisition, as demonstrated by Ellis (2002), “frequency should also play a key role in the development of materials and in the choices that teachers make in language classrooms. With the recent availability of comprehensive frequency-based grammatical descriptions, such integration of pedagogy and research has become feasible” (Biber & Reppen, 2002, p. 206-207).

7.4 Applications for Pedagogical Materials

7.4.1 Survey

Given the discussions of pedagogical materials and corpus findings in them (or lack thereof), as well as the ideas proposed throughout this thesis, this section presents a survey of six popular academic writing textbooks used in the EAP setting in North American IEPs, particularly with reference to the use of models and text types and coverage of grammar and the noun phrase. The particular textbooks were chosen by emailing Higher Education Representatives from four current, mainstream ELT publishers in the North American EAP context serving IEPs—Cambridge University Press, Pearson Longman Education, Heinle Cengage Learning, and the University of Michigan Press—requesting the names of their top-

selling academic writing textbooks for this setting. The University of Michigan Press did not have any academic writing textbooks appropriate for the IEP setting, as their books focus mainly on graduate level writing (See Swales & Feak, 2009, as an example) (Sippel, 2010). The other three publishers, however, provided the names of their top two selling textbooks (Brezinsky, 2010; Seal, 2010; Driscoll, 2010) and also agreed to send review copies of the identified textbooks.³ Here, in section 7.4.1, I provide an overview of the textbooks and a review of the grammar taught in them. In section 7.4.2 I discuss the results of the review and make further recommendations for pedagogical applications.

Each of the top selling academic writing textbooks for the IEP setting (for all publishers) were the upper-level texts from multi-level writing series. Specifically, the six textbooks under review include *Writers at Work: The Essay* (Zemach & Stafford-Yilmaz, 2008); *Writers at Work: The Short Composition* (Strauch, 2005); *Introduction to Academic Writing* (Oshima & Hogue, 2007); *Writing Academic English* (Oshima & Hogue, 2006); *Great Essays* (Folse, Muchmore-Voloun, & Vestri Solomon, 2010); and *Greater Essays* (Folse & Pugh, 2010). An overview of the textbooks is provided in Table 7.1.

Table 7.1: Overview of IEP Writing Textbooks

Textbook	Publisher	From the Back Copy
<i>Writers at Work: The Essay (WW4)</i>	Cambridge University Press	“ <i>Writers at Work: The Essay</i> is the third book in a four-book series that provides students with a solid foundation in writing skills. The book adopts a process approach in which students draw from their personal experience to compose coherent and accurate multi-paragraph compositions. Students also learn how to write about works of fiction and nonfiction by summarizing and citing sources.”
<i>Writers at Work: The Short Composition (WW3)</i>	Cambridge University Press	“ <i>Writers at Work: The Short Composition</i> is the fourth book in a four-book series that provides students with a solid foundation in writing skills.

³ A special thanks to Bernard Seal, Christian Lorentzo, Pietro Alogni, and Sarah-Jane Platt for sending copies of the books.

		The book adopts a process approach in which students use critical thinking and personal experience to compose well-structured and accurate essays. Students also learn how to incorporate source material into their essays and to write timed essays under exam conditions.”
<i>Introduction to Academic Writing (IAW)</i>	Pearson Longman Education	“The Third Edition of <i>Introduction to Academic Writing</i> , by Alice Oshima and Ann Hogue, continues in the tradition of helping students to master the standard organizational patterns of the paragraph and the basic concepts of essay writing. The text's time-proven approach integrates the study of rhetorical patterns and the writing process with extensive practice in sentence structure and mechanics.”
<i>Writing Academic English (WAE)</i>	Pearson Longman Education	“Now in its long-awaited fourth edition, <i>Writing Academic English</i> is the essential writing text to prepare high-intermediate to advanced college or college-bound English language learners for academic success.”
<i>Great Essays (GE)</i>	Heinle Cengage Learning	“ <i>Great Essays</i> is the fourth book in a five book series that uses a wide variety of writing models in carefully selected rhetorical styles that provide practice in working with the writing process to develop a final piece of writing. The book uses clear explanations and extensive practical activities to help students write effective essays.”
<i>Greater Essays (GRE)</i>	Heinle Cengage Learning	“ <i>Greater Essays</i> is the fifth book in a five book series that uses a wide variety of writing models in carefully selected rhetorical styles that provide practice in working with the writing process to develop a final piece of writing. The book uses more challenging models and assignments to give students opportunities to further develop their essay writing and vocabulary skills.”

7.4.1.1 “Writers at Work”

Writers at Work is a four-book series from high beginning to high intermediate levels that adopts a process approach to teaching writing in order to enable students to, upon completion of the series, “tackle academic essay writing” (Cambridge University Press, 2010). Specifically, *Writers at Work: The Essay (WW4)* aims to provide training in how to write different genres of essays common at the post-secondary level; chapters in the text are divided

according to text type—explanatory essays, problem-solution essays, comparison-contrast essays, persuasive essays, responding to a reading, and timed essays. *Writers at Work: The Short Composition (WW3)* also aims to provide instruction in different essay genres, but the chapters are classified according to language point as well as text type: writing about a poem, narrating a personal experience, writing a summary, responding to non-fiction, critiquing fiction, interpreting quotations and proverbs, then providing examples, supplying reasons, supporting with parallel points.

Both texts use a five part chapter structure: Getting Started, Preparing the First Draft, Revising Your Writing, Editing Your Writing, and Following Up. Both texts also make use of model essays, provide exercises for noticing key organizational and linguistic features, and specifically focus on selected aspects of grammar in the Editing Your Writing part of the chapter structure; more specifically, in the Preparing the First Draft part of the chapter structure in *WW4*, students “study language structures that are likely to occur in the type of essay featured in the chapter” (p. xii). *WW4* does not, however, present any material related to the noun phrase.

WW3 provides specific instruction on article usage: article before a singular count noun, no article before non-count nouns in general statements, indefinite article used for first mention of singular count noun, definite article used for reference, and definite articles in specific contexts. Figures 7.1-7.4 illustrate the presentation of articles from Chapter 6 of *WW3*, “Interpreting Quotations and Proverbs,” Editing Your Writing section. (Exercises to review the material have not been included as the practice exercises themselves are not under review.)

Figure 7.1 Articles with Count Nouns (*Writers at Work: The Short Composition*, Strauch, 2005, p. 124)

IV EDITING YOUR WRITING

A Edit for correct use of articles

ARTICLES WITH COUNT NOUNS

Use the articles *a*, *an*, and *the* (or other determiner) before a singular count noun.

- a book, a truck, a bridge, a cup of coffee, a personal experience
- an example, an experience, an accident
- the box, the proverb, the edge, the darkness

Figure 7.2 Articles with Non-count Nouns (*Writers at Work: The Short Composition*, Strauch, 2005, p. 124)

NONCOUNT NOUNS

Do not use *the* before a noncount noun when the noun is being used in a general statement.

Incorrect	Correct
The music calms the soul.	Music calms the soul.
The honesty is the best policy.	Honesty is the best policy.
The love cures all.	Love cures all.

Figure 7.3 Indefinite and Definite Articles (*Writers at Work: The Short Composition*, Strauch, 2005, p. 125)

INDEFINITE AND DEFINITE ARTICLES

The indefinite article *a* is commonly used before the first mention of a singular count noun.

I was watching a man cross the street when suddenly a fire engine came screaming through the intersection . . .

However, in later uses of the same noun, the definite article *the* is normally used because the reference has become definite, or known.

. . . The man jumped onto the curb to avoid the fire engine.

Figure 7.4 Definite Articles in Specific Contexts (*Writers at Work: The Short Composition*, Strauch, 2005, p. 126)

DEFINITE ARTICLES IN SPECIFIC CONTEXTS

In certain contexts, some nouns take the definite article.

- In the context of a person's house
the door, the floor, the kitchen counter, the edge of the counter
- In the context of the public environment
the bank, the post office, the mail carrier, the library, the street
- In the context of nature
the moon, the sun, the beach, the desert, the stars

A definite article is used before a count noun when the noun is followed by specific information that modifies it.

Last night I visited an unusual house.

Last night I visited the unusual house of my father's friend.

Today I learned an English proverb.

Today I learned the English proverb "Haste makes waste."

As reported in Chapter 6, the definite article *the* is used 85% of the time in scholarly writing; this figure would seemingly make it especially important for grammar instruction dealing with academic writing and the emphasis placed on its use here noteworthy. However, *the* is used only slightly more than one fourth of the time as the determiner in a noun phrase in the NAFWiC, and if one is using freshman composition as a model for EAP writing in the IEP, the need for such instruction is not as great. This does not preclude instruction on the definite article, however. Even though *the* is used only a quarter of the time in the NAFWiC compared to scholarly writing, students must still be aware of when it is appropriate (i.e. that 25%). What is more important to note here is not what is included, but what is excluded. The definite article is only one choice for determiner—of three possible—in a definite noun phrase; possessive and demonstrative determiners should also be introduced (Biber, et al, 1999, p. 270).

Figure 7.4 rightly notes that a definite article is used in cataphoric reference (the referent is forward); and while this is the most common type of reference pattern for a definite noun phrase, as discussed in Chapter 6 section 6.3.1.1, it is used so only 40% of the time; definite noun phrases are also used in anaphoric reference 25% of the time. Figure 7.3 does note this function (naming a previous referent), but does so in isolation of cataphoric reference. Both statements are, as discussed in section 7.2.2, falsely portrayed as absolute (Hyland, 1994) and may appear confusing to students. Is it the first or the second? Acknowledgement that it is both, with a slight tendency for cataphoric reference, would be useful.

7.4.1.2 Longman Academic Writing Series

The Longman Academic Writing series consists of four books from beginning to advanced levels to “give students the pedagogical support to quickly improve their writing as well as provide them with tools for academic success” (Pearson Longman, 2010).

Introduction to Academic Writing (IAW) is the third book of the series, aimed at intermediate level writers, and aims to teach rhetoric and sentence structure while guiding students through the process approach to writing. The first eight chapters of *IAW* teach paragraphs, with the last two chapters introducing the essay; chapters either focus specifically on text type (narrative, descriptive, process, comparison/contrast, definition, and opinion) or paragraph construction (format, structure, logical division of ideas, and organization). *Writing Academic English (WAE)*, the fourth and highest level book of the series, also describes itself as a comprehensive rhetoric and sentence structure textbook and is divided into three parts: focusing on a quick review of paragraph writing and summarizing in the first part of the book, chapters of text types process, cause/effect, comparison/contract, and argumentative essays in

the second part of the book, and sentence structure with emphasis on subordinated structures in the third part of the book.

Most chapters in *IAW* are divided into three sections—Organization, Sentence Structure, and Writing—with most chapters teaching sentence structure, starting with simple sentences and progressing through compound and complex sentences, as well as including instruction in punctuation. *IAW* also makes widespread use of student writing models, for noticing in both the Organization (e.g. concluding sentence) and Sentence Structure (e.g. descriptive word) sections. *WAE* does not have a specifically outlined consistent chapter structure, though each chapter in parts 1 and 2 of the text generally begins with presentation of an instructional topic (e.g. the three parts of a paragraph, unity, facts versus opinions); provides examples, exercises, and a review; then moves to writing practice followed by an Applying What You Have Learned activity (which involves noticing of key ideas in model writing).

The Criterion online writing evaluation program, where students electronically submit writing for software to identify grammatical errors, is currently available for *IAW* and *WAE*; additionally, of the six appendices in *IAW* and *WAE*, three in *IAW* deal directly with grammar instruction and two in *WAE* do so. Because the Criterion online writing evaluation program and the appendices are supplementary materials are not part of the regular chapter structure, the grammar covered there is not included below.

WAE does not present any material related to the noun phrase; it does, however, devote an entire chapter (Chapter 12) to the noun clause (this will be discussed more in section 7.4.2.3). *IAW* provides specific instruction on appositives. As reported in Chapter 6, appositives occur 7% of the time as post-modifiers in the noun phrase in scholarly writing, but

only just over 1% in the NAFWiC. Figures 7.5-7.6 illustrate the presentation of appositives from Chapter 8 in *IAW*, “Definition Paragraphs.”

Figure 7.5 Appositives (*Introduction to Academic Writing*, Oshima & Hogue, 2007, p. 30)

Appositives	<p>Appositives are nouns or noun phrases that rename a preceding noun or noun phrase.</p> <ol style="list-style-type: none">1. They set up a special table on which they display seven foods with names beginning with the letter s in Farsi, the language of Iran.2. There people celebrated the end of the farming season by lighting bonfires, large outdoor fires, to keep away bad spirits who might appear in the night. <p>In sentence 1, <i>Farsi</i> and <i>the language of Iran</i> are the same thing. In sentence 2, <i>bonfires</i> and <i>large outdoor fires</i> are the same thing. <i>The language of Iran</i> and <i>large outdoor fires</i> are appositives.</p> <p>Appositives are very useful in writing definitions because they give the reader more information about your topic concisely (without a lot of extra words).</p> <p>Fudge, a delicious chocolate candy, was invented as a result of a cooking mistake.</p> <p>Appositives can be necessary information or extra information.¹ Consider this sentence:</p> <p>My friend Tim got married last week.</p> <p>In this sentence, <i>Tim</i> is an appositive because <i>Tim</i> and <i>my friend</i> are the same person. <i>Tim</i> is a necessary appositive because it is necessary to identify which friend got married. If we omit the word <i>Tim</i>, we don't know which friend got married.</p> <p>Now consider this sentence:</p> <p>Tim, my friend, got married last week.</p> <p>In this sentence, the appositive is <i>my friend</i>. It is extra information because the name <i>Tim</i> already identifies the person who got married. If we omit <i>my friend</i>, we still know who got married. The fact that he is the writer's friend is not necessary to identify him. It is merely extra information.</p> <p>If there is only one of an item, it is unnecessary to identify it further, so appositives of one-of-a-kind items are always extra information. For example, Earth has only one moon, so any appositive of <i>the Moon</i> in a sentence would be extra information. Similarly, adjectives such as <i>tallest</i>, <i>strongest</i>, <i>oldest</i>, and <i>most interesting</i> automatically make the following noun one of a kind.</p> <p>The Moon, Earth's only natural satellite, orbits Earth about once a month.</p> <p>The highest mountain in North America, Mount McKinley, is in Alaska.</p> <p>The appositive <i>Mount McKinley</i> is extra information because it follows <i>the highest</i>, and there can be only one “highest.”</p> <p>I have three sons. My son Carlos looks like me.</p> <p><i>Carlos</i> is necessary information because there are three sons, and the name <i>Carlos</i> is necessary to identify which son.</p>
--------------------	---

Figure 7.6 More on Appositives (*Introduction to Academic Writing*, Oshima & Hogue, 2007, p. 131)

<p>My youngest son, Javier, looks like his father.</p> <p>The appositive <i>Javier</i> is extra information because there can be only one youngest son.</p> <p>Comma Rule</p> <p>Use commas to separate an extra information appositive from the rest of the sentence. Do not use commas with necessary appositives.</p>
--

Figure 7.5 introduces the students to appositives as “nouns or noun phrases” and talks about their function in relation to a “noun or noun phrase.” Although the term “post-modifier” is not used, the information in Figure 7.5 provides students with important information. The only information that might be added is that appositives are not used very often. This does not mean that appositives should not be taught—indeed, exposure to appositives may serve useful in student writing—but students should be given a perspective of how frequently they can expect to see, or might use, this structure. As Figures 7.5 and 7.6 show, appositives can be used to identify a person or an object. Investigating the function of the appositive in the NAFWiC would be useful in evaluating this information. We do know that in scholarly writing, appositives typically refer to a proper noun or technical name—rarely human reference—and often provide an explanation to a technical reference or introduce an acronym. Noting this additional function could also be useful for learners.

7.4.1.3 “Great Writing”

Great Writing is a five-book series from basic to advanced levels that uses a variety of writing models in selected rhetorical styles to provide practice in the writing process. *Great Essays (GE)* is the fourth book in the series and aims to “provide introductory instruction and extensive practical exercises and activities in essay writing at the high-intermediate and advanced levels” (p. x). The book opens with a unit that presents the overall organization of

an essay, including suggestions for writing an introduction (including how to write a good hook and solid thesis statement), with the remaining four units of the book presenting different text types: narrative, comparison, cause-effect, and argumentative. *Greater Essays* (*GRE*) is the fifth book in the series and aims to give students opportunities to develop their essay writing skills as well as their overall language skills. The first unit in *GRE* also provides a general overview of essays, with the remaining five units discussing in detail the text types process analysis, comparison, cause-effect, argumentative, and narrative.

Units in both *GE* and *GRE* contain the following 11 features: Example Essays, Writer's Notes, Language Focus, Building Better Vocabulary, Building Better Sentences, Completing an Outline, Completing a Sample Essay, Analyzing an Essay, Topics for Writing, Timed Writing, and Peer Editing; *GRE* contains an additional feature, Building Better Grammar. Through these features, the books present models of good academic essays, offer direct vocabulary instruction—with a focus on collocation, address sentence combination skills, and provide multiple opportunities to practice timed writing. Though both books introduce and provide exercises in the writing process, in *GE*, “the focus is slightly more on the final written product” (p. xi); in *GRE* the focus is on the revision process, particularly helping students become “better writers by learning to become better editors of their own and their peers’ essays” (p. xi).

In *GE* and *GRE*, the Building Better Sentences and Building Better Grammar activities are referenced as part of the instruction within the unit, but the actual exercises are located in the appendix; because the feature is integrated as part of the regular unit structure, the grammar covered in Building Better Sentences/Grammar is included as part of the regular unit and addressed below. Both texts also contain a Brief Writer's Handbook with activities, but in *GE* these activities focus on “Understanding the Writing Process” while in *GRE* the

activities solely address grammar. Because the handbooks are supplementary and not part of the regular unit structure, the grammar covered there is not included below.

Grammar instruction in *GE* primarily takes place through the Building Better Sentences unit feature. Each unit contains three to seven of these prompts which instruct students to go to the appendix for further practice in correct and varied sentence structure. *GRE* also utilizes the Building Better Sentences unit feature, though each chapter contains only one prompt. The primary grammar instruction in *GRE* occurs through the Building Better Grammar exercises which aim to help refine grammar skills and help students become better editors. The Building Better Grammar exercise which deals with articles is illustrated in Figure 7.7 below.

Figure 7.7 Articles (*Greater Essays*, Folse & Pugh, 2010, p. 17)

Grammar Topic 1.4
Confusing Words: Articles A/An

A and an are articles that come before nouns.

a cat an elephant

Sometimes an adjective comes between a or an and the noun it modifies.

a black cat an interesting cat

Use an before words that begin with a vowel sound.

an umbrella an open door an honest man

If a word begins with a vowel but not a vowel sound, do not use an. Instead, use a.

a university a uniform

The inclusion of this information in an advanced level writing book is a bit surprising. While articles certainly can be confusing words, whether to use *a* or *an* would seem to be the least of these complications, especially for the advanced level student at whom this book is aimed. Any number of useful instructional points stemming from the findings in Chapter 6 (such as those mentioned above in regards to *WW3*) could accompany this information.

The materials from Appendix 1, “Building Better Sentences,” are illustrated in Figures 7.8-7.9.

Figure 7.8 Building Better Sentences (*Great Essays*, Folse, Muchmore-Voloun, & Vestri Solomon, 2010, p. 161; *Greater Essays*, Folse & Pugh, 2010, p. 202).



Building Better Sentences

Being a good writer involves many skills, such as being able to use correct grammar, vary vocabulary usage, and state ideas concisely. Some student writers like to keep their sentences simple because they feel that if they create longer and more complicated sentences, they are more likely to make mistakes. However, writing short, choppy sentences one after the other is not considered appropriate in academic writing. Study these examples:

The time was yesterday.
It was afternoon.
There was a storm.
The storm was strong.
The movement of the storm was quick.
The storm moved towards the coast.
The coast was in North Carolina.

Notice that every sentence has an important piece of information. A good writer would not write all these sentences separately. Instead, the most important information from each sentence can be used to create one longer, coherent sentence.

Read the sentences again below and notice that the important information has been circled.

The time was yesterday.
It was afternoon.
There was a storm.
The storm was strong.
The movement of the storm was quick.
The storm moved towards the coast.
The coast was in North Carolina.

Here are some strategies for taking the circled information and creating a new sentence.

1. Create time phrases to introduce or end a sentence: *yesterday + afternoon*
2. Find the key noun: *storm*
3. Find key adjectives: *strong*
4. Create noun phrases: *a strong + storm*
5. Change word forms: *movement = move; quick = quickly*
moved + quickly
6. Create prepositional phrases: *towards the coast*
towards the coast (of North Carolina)
or
towards the North Carolina coast

Now read this improved, longer sentence:

Yesterday afternoon, a strong storm moved quickly towards the North Carolina coast.

Figure 7.9 More on Building Better Sentences (*Great Essays*, Folse, Muchmore-Voloun, & Vestri Solomon, 2010, p. 162; *Greater Essays*, Folse & Pugh, 2010, p. 203).

Here are some more strategies for building better sentences:

7. Use coordinating conjunctions (*and, but, or, nor, yet, for, so*) to connect two sets of ideas.
8. Use subordinating conjunctions, such as *after, while, since, and because*, to connect related ideas.
9. Use clauses with relative pronouns, such as *who, which, that, and whose*, to describe or define a noun or noun phrase.
10. Use pronouns to refer to previously mentioned information.
11. Use possessive adjectives and pronouns, such as *my, her, his, ours, and theirs*.

Study the following example.

Susan went somewhere. That place was the mall. Susan wanted to buy new shoes.
The shoes were for Susan's mother.

Now read the improved, longer sentence:

Susan went to the mall because she wanted to buy new shoes for her mother.

The number 4 strategy, “create noun phrases” is obviously good advice for academic writers in the IEP setting. While more information could be given on this subject, the Building Better Sentences instructions shown here are the only mention of creating or using a “noun phrase” in all the textbooks under review. Strategies 2 and 3 (“find the key noun” and “find the key adjective”, respectively) are also good strategies for creating noun phrases. To provide more focus on noun phrases, Strategies 6 and 9 (“create prepositional phrases” and “use clauses with relative pronouns”, respectively) could also be targeted toward the creation of noun phrases; as discussed earlier, more than 80% of noun phrase post-modifiers in the NAFWiC are either prepositional phrases or relative clauses. Strategy 10 (“use pronouns to refer to previously mentioned information”), however, could be removed from this otherwise useful list. As reported in Chapter 6, nouns comprise nearly 50% of all lexical words used in the NAFWiC; repetition of the key noun is important for clarity in academic writing (rather than the substitution of pronouns).

7.4.1.4 Summary

Though all the textbooks under review deal with presenting both grammar and writing, each series handles the integration of the two skills differently. Table 7.2 displays the specific grammatical structures presented in each textbook. A brief description of the treatment of grammar is provided here, while section 7.4.2.3 considers the coverage from a critical perspective. The *Writers at Work* Series includes all grammar instruction directly with the unit features, but the focus is on writing practice, supplementing writing practice with grammar presentation. The Longman Academic Writing series seems to place equal value on both writing and grammar, including grammar instruction in regular unit features as well as providing a plethora of additional grammar activities; *IAW* and *WAE* have both a wider variety and a greater depth to the grammar presented than in the other series. Grammar instruction in *GE* primarily takes place through the Building Better Sentences unit feature where exercises are contained in the appendixes, putting grammar on the periphery. *GRE* also utilizes the Building Better Sentences unit feature, but the primary grammar instruction in this text occurs through the Building Better Grammar exercises. The Building Better Sentences unit featured in both *GE* and *GRE* is extensive reference to additional sentence combination exercises which focus on 11 strategies for making longer and more complicated sentences. It is worth noting, perhaps, that the *Great Writing* Series also differs markedly from the other series under review in its considerable attention to vocabulary development.

Table 7.2: Grammar Features in IEP Writing Textbooks

	<i>WW3</i>	<i>WW4</i>	<i>IAW</i>	<i>WAE</i>	<i>GE</i>	<i>GRE</i>
Subject-Verb Agreement	X		X			X
Simple past/past progressive action verbs	X				X	
<i>Subordinating conjunctions</i>	X	X	X	X	X	X
<i>Coordinating conjunctions</i>	X	X	X	X	X	X
Correlative conjunctions				X		
Run-on sentences	X	X	X	X		
‘Stringy’ sentences	X			X		

Independent clauses	X	X		X		
Dependent clauses	X	X		X		
Sentence fragments	X		X			X
<i>Simple sentences</i>	X	X	X	X	X	X
<i>Compound sentences</i>	X	X	X	X	X	X
<i>Complex sentences</i>	X	X	X	X	X	X
Compound-complex sentences				X		
Articles	X					X
The 'literary' present tense	X					
Punctuation			X			X
Modals	X	X			X	X
"Because" and "therefore" transitions		X				
Hedging (qualifying phrases, modals, modifiers, adverbs)		X				
Conditional sentences		X				
Academic language for comparing/contrasting (e.g. "in the same way", "similarly", "although")		X	X	X	X	
Comparative forms		X				X
Argumentative language (e.g. "some...however", "while...also")		X				
Reporting verbs		X		X		
Repeating key nouns				X		
Transition signals for coherence			X	X		
Noun clauses ("that", "if"/"whether" & question)				X		
Noun phrases					X	X
Adverb clauses (time, place, distance, frequency, manner, reason, result, purpose, contrast, conditional)				X	X	X
Adjective clauses (relative, possessive, time, place, participial phrases)			X	X		
Present perfect (versus simple past)		X				X
Appositives			X			
Relative pronoun clauses					X	X
Prepositional phrases /preposition combinations					X	X
Connectors for cause-effect essays (e.g. "therefore", "as a result of")					X	
Connectors for chronological order (e.g. "first", "next", "then")					X	
Using "this" and "these"						X
Parallelism						X
"Used to" + verb						X
Non-count nouns						X
Verb tense					X	X

Homonyms						X
Pronouns					X	X
Possessive adjectives and pronouns					X	X
Collocations					X	X
Word forms					X	X

Italics indicates presence in all textbooks; **bold** indicates relation to the noun phrase

Only five grammar points are covered in all the textbooks, and each of these points revolves around the creation of longer sentences (sentence types and coordinating and subordinating conjunctions). Additional overlap, however, can be seen among the various levels of the textbooks; *WW3* and *IAW* (Oshima & Hogue, 2007), two of the lower levels presented here, both deal with sentence fragments and subject-verb agreement; *WW4* (Zemach & Stafford-Yilmaz, 2008) and *WAE* (Oshima & Hogue, 2006), two of the higher levels presented here, and the highest levels of their respective series, both deal with reporting verbs. Though there is some consensus among the textbooks under review as to what grammar should be presented in accordance with academic writing, the consensus is not very strong; in addition, the noun phrase is not a part of that consensus.

While all the textbooks under review provide a sizeable amount of grammar instruction for academic writing, the noun phrase is only a tiny portion of that which is covered in the regular class materials: *IAW* (p. 194-209) provides specific instruction on appositives, and *WW3* (p. 124-127) provides specific instruction on article usage. The Great Writing Series does specifically instruct students to “create noun phrases” in the Building Better Sentences unit feature, but there are no indications in the textbook of what this means or looks like. It may be assumed that students are familiar with this concept but perhaps need practice. One Building Better Grammar exercise in *GE* (p. 17) looks at basic article usage.

Clearly more focus on the noun phrase is needed in EAP writing textbooks for the IEP. As demonstrated above, few aspects of the noun phrase, heavily employed in academic

writing at all levels, are addressed in the textbooks under review, and when those aspects are addressed, additional or relevant details learned from the corpus have been omitted. The next section provides more insight into the grammatical structures which appear in these texts as well as recommendations based on the proposed model from Chapter 2 and schematic representation from Chapter 6.

7.4.2 Recommendations

7.4.2.1 Models

An important argument in this thesis is that writing models are valuable for instruction, but that the model informing EAP writing in the IEP should be based on student writing—both the language structures and functions which comprise the “teaching points” (such as from the schematic representation) in a textbook and the sample writing shared with learners. Although the textbooks under review do not seem to base teaching points off of a schematic representation (of either expert or scholarly writing), one aspect which does merit attention is their widespread use of student writing as a model; all of the textbooks claim to use successful student writing models throughout the presentation and exercises in the textbooks, and as Jordan (1997) advocates, a provided model and undertaking of various exercises to draw attention to important features, followed by the production of a similar or parallel text, is useful for academic writing instruction.

7.4.2.2 Text Types

As discussed in Chapters 2 and 4, repeated assignments and use of text types for students, including those in freshman composition, have led to a “student writing genre;” some scholars do consider “student writing” as inauthentic writing with no real communicative purpose, but this is an unfounded claim (students are obviously writing to an audience and have a purpose for doing so!). Another noteworthy feature of the textbooks

under review is their use of text types to organize chapters and units, with each textbook covering at least two of the text types contained in the NAFWiC. Coverage of text types in each of the textbooks under review is presented in Table 7.3. More than half of the text types represented in the NAFWiC were assigned by at least one textbook here. At least one textbook assigns persuasive writing, personal narrative, summary synthesis, rhetorical analysis, profiling, remembering an event, and proposing a solution. Perhaps not surprisingly, the two text types that received the most attention from the textbooks are the personal narrative and the persuasive essay; all the textbooks assigned the personal narrative and five of the textbooks (all but one) a persuasive essay. The persuasive essay is one of two text types from the NAFWiC assigned at the majority of the institutions represented in it. The other text type assigned by the majority of institutions is the proposing a solution; only one textbook assigned a proposing a solution essay. None of the textbooks, however, address expository writing, film analysis, literary research paper, or thematic or character analysis.

Table 7.3 Text Types in the NAFWiC and Academic Writing Textbooks

<i>NAFWiC</i>	<i>WWSC</i>	<i>WWE</i>	<i>IAW</i>	<i>WAE</i>	<i>GE</i>	<i>GRE</i>
Expository Essay						
Persuasive Essay		X	X	X	X	X
Personal narrative	X	X	X	X	X	X
Summary synthesis	X		X			
Rhetorical analysis		X	X			
Remembering an event			X			
Profile			X			
Proposing a solution		X				
Film Analysis						
Literary Research Paper						
Thematic Analysis						
Character Analysis						

There were two text types assigned in all or a majority of the textbooks that are not in the NAFWiC. Two-thirds of the textbooks (the Longman Academic Writing Series and the Great Writing Series) assigned process writing, though this text type is not found in the

NAFWiC. Given the diversity of the course, it may be that freshman composition instructors do assign process essays; however, that text type is not found here nor is it listed in Ferris (2011) or Wardle (2009) as among the most common. All but one of the textbooks assigned a “comparison/contrast” essay. Although the persuasive essay and proposing a solution text types require the discussion of one topic from two angles (thus requiring a kind of compare and contrast), the assignments in the textbooks focused on comparing and contrasting two topics; the difference is the first involves persuasion and the second description. One more point of comparison to make between the textbooks and the text types involves writing from sources. Nine of the text types from the NAFWiC (75%) require the use of at least one source; five require the use of multiple sources. Only two of the textbooks here actually require an outside source (although the Great Writing series does include an information sheet on plagiarism in the Appendix, it is not referenced or required as part of the assignments); text types which require writing from sources would certainly be a worthy addition for the textbooks.

7.4.2.3 Grammar

Section 7.2.2 considered factors which go into the design and development of textbooks. One factor an author must consider in developing a textbook (or an instructor must consider in adopting one) is his or her definition of the skill being taught. Of relevance here, how you define writing or grammar affects your approach to it. Chapters 2 and 3 explored the difficulties in defining “good writing,” as Hjortshoj (2001, p. 33) noted, “features of good writing often vary from one situation to another” and the “field of study often structures those expectations.” Although the issue of defining grammar has not been considered here, I adopt Byrd’s (1994, p. 246) definition that grammar is “structures in use in a particular context.” For writing materials in the IEP, the field of study can be freshman composition which all

North American university-bound EAP students, indeed all undergraduate students, must confront in order to obtain the higher education degree they seek; grammar would, then, be the structures used in freshman comp (or, for more general applications, scholarly writing). In the teaching of writing, an author (an instructor) must also consider their views on the role of grammar instruction in writing.

Chapter 2 discussed the desire of EAP writing methods and materials to focus on both product (form) and process. Form, however, does not mean rote or decontextualized; it considers both meaning and structure. As noted in Chapter 3 section 3.4.1, because learner writers are still developing proficiency in the L2, form is just as demanding as content; research has shown that it is in the best interest of L2 writers to attend to language issues consistently throughout the draft process (Frodesen, 2003, p. 145). Leki (2007, p. 83), who conducted a five-year quantitative follow-up study of students at American universities who had completed EAP writing courses, found that, in interviews about the usefulness of those classes, students invariably answered “English required grammar and these classes provided grammar.” This certainly supports the perceptions of EAP writing instructors who feel they “have the responsibility for creating opportunities in which learners can come to a fuller understanding of...the linguistic choices which have to be made in order to produce allowable contributions” in their target learning environment (Tribble, 2002, p. 131-132); or the EAP practitioners who are concerned with understanding “features of academic writing which could be passed on to students” (Hewings, 2004, p. 131-132). In fact, Fodesen and Holton (2003, p. 157) point out that “the question is not *whether* we should ‘teach’ grammar but *how best* to do it.” The attention to grammar by the textbooks under review also stresses the importance of grammar to EAP writing.

How to teach grammar, though, is not quite as consensual. Many EAP scholars advocate for noticing or consciousness-raising activities to help students become aware of target forms (Biber, 1995; Hyland & Milton, 1997; Hyland, 2002b; Frodesen & Holt, 2003; Hinkel, 2004b) and for the grammatical features that cluster together in target texts to be considered in making decisions about which structures to teach (Frodesen & Holt, 2003; Coxhead & Byrd, 2007). Though the textbooks under review do use student models to outline structure of text types, noticing use of the grammatical structures is not a highlighted activity [with the exception of the Longman Academic Writing series which does have students notice certain features (e.g. conjunctions) in the student models]. And, as discussed in section 7.4.1, there is not a strong consensus on what grammar structures should be highlighted for EAP writers in the IEP, and there is certainly little concern for the noun phrase. At all levels of instruction, even the advanced IEP level, grammar instruction in published writing textbooks seems to focus primarily on sentence construction in order to make longer sentences, with this focus involving the use of coordinating and subordinating conjunctions to do so. Hinkel (2003, p. 276) demonstrated that non-native speakers tend to employ excessively simple syntactic and lexical constructions, which often prove to be a handicap in rated writing, but that “subordinate clauses...lend a degree of sophistication to a text.” Thus, this focus seems justifiable. But, there are two problems here. Firstly, many of the textbooks here focus on a “deficit” model of grammar teaching; address weaknesses, fix what is wrong. Yet, as Evans and Green (2007, p. 12) note, “remediation or general language proficiency is unlikely (by itself) to help students meet the new challenges of writing in the academy; nor would such a programme be especially motivating as students would perceive it as ‘more of the same.’” Furthermore, there is a “false” impression that clausal structures are more difficult and useful for students; for example, an entire chapter in *WAE* focused on noun

clauses. In fact, however, as discussed in Chapter 6, phrasal structures are more characteristic of academic writing at both expert and novice levels, and can be quite difficult. Biber and Gray (2010) and Biber, Gray, and Poonpon (2011, p.7) claim that “the most important structural characteristic of academic written discourse: the reliance on phrasal rather than clausal elaboration.” And because students already have a tendency to employ language more characteristic of conversation (as discussed in Chapters 5 and 6), a focus only on clauses may complicate language matters for them. Perhaps more importantly, as discussed in Chapter 3 section 3.5.6, writers currently on level 1 of the staircase model employ fewer lexical and syntactic modifiers of nouns and fewer modifying prepositional phrases. These factors intensify the need for textbooks to focus on the noun phrase.

7.4.2.4 The Noun Phrase

While several suggestions were made throughout section 7.4.1 for including grammar more in alignment with successful freshman composition writing, all materials which aim to teach EAP writing in the IEP setting should consider the inclusion of at least two key language features: the use of adjective pre-modifiers (used approximately 85% of the time in noun phrases in the NAFWiC) and prepositional phrase post-modifiers (used approximately 65% of the time in noun phrases in the NAFWiC). The addition of these language features is a solid starting point for addressing the grammar these writers will need in their academic writing in the near future. It should also be noted that these two features, among the most utilized in successful freshman composition writing as represented in the NAFWiC, can also help with creating longer, more complicated sentences, a heavy grammatical focus of the current academic writing textbooks, and thus not force a potential compromise by the authors or publishers.

In addition to highlighting adjective pre-modifiers and *of*-phrase post-modifiers, the breakdown of the noun phrase into teaching aspects of modification—pre-modifiers and post-modifiers—would be a practical method for presenting the noun phrase to students. Textbooks at the high-intermediate to advanced levels, such as those reviewed here, could efficiently add information such as that exemplified in Figures 7.10 to 7.13 to their grammar instruction; the materials are specifically based on the comparative results presented in Chapter 6 and in response to the review conducted above. While I would use materials presented in these figures (and have used materials very similar to them) in my own classroom, not all teachers would be comfortable presenting such technical information; nevertheless, students need access to this type of information, and materials writers could easily use the figures and their information as sources for more traditional classroom materials. Use of the materials in class is discussed below.

Figure 7.10 Introduction of the Noun Phrase

Introduction to the Noun Phrase

One of the most important grammatical features of successful first year undergraduate writing is its extensive use of nouns and noun phrases. In fact, nearly 50% of all content words—that is nouns, verbs, adjectives and adverbs—are nouns!

A noun phrase is created when more information is given about a noun. This information is called a modifier. There are three types of noun phrases in successful first year undergraduate writing:

1. Those with pre-modifiers (where more information is given before the noun).
2. Those with post-modifiers (where more information is given after the noun).
3. Those with both a pre-modifier and a post-modifier (where more information is given before *and* after the noun).

- Nearly 80% of nouns have at least one modifier.

Figure 7.11 Presentation of Noun Phrases with Pre-modifiers

Noun phrases with pre-modifiers

A noun that has additional information is modified. The additional information is given with a modifier. When a modifier is before the noun, this is a pre-modifier.

the American man is a noun phrase.

The noun is *man* and the additional information about the man is *the American*. Because this information is given before the noun *man*, it is a pre-modifier.

More examples of noun phrases with pre-modifiers:

her Puerto Rican culture

a pretty white baby

the Italian brothers' statement

- Nearly 35% of nouns have a pre-modifier.

The most common pre-modifiers in successful first year undergraduate writing are adjectives and nouns.

- An adjective is used more than 85% of the time in as a noun pre-modifier.
- A noun can also be used as a pre-modifier, but is not as common (less than 15% of the time).

the American man → *the* + adjective + noun

her Puerto Rican culture → possessive pronoun + adjective + noun

Sometimes more than one pre-modifier can be used.

a pretty white baby → *a* + adjective + adjective + noun

the Italian brothers' statement → *the* + adjective + noun + noun

Figure 7.12 Presentation of Noun Phrases with Post-modifiers

Noun phrases with post-modifiers

A noun that has additional information is modified. The additional information is given with a modifier. When a modifier is after the noun, this is a post-modifier.

issues that are debated in the news is a noun phrase.

The noun is *issues* and the additional information about the issues is *that are debated in the news*. Because this information is given after the noun *issues*, it is a post-modifier.

More examples of noun phrases with post-modifiers:

people who work hard all their lives

facets of the university system

mistakes in life

- 20% of nouns have a post-modifier.

The most common post-modifiers in successful first year undergraduate writing are prepositional phrases and relative clauses.

- A prepositional phrase is used approximately 65% of the time in post-modifiers.
 - Nearly half of the time (30%), it is a prepositional phrase beginning with *of*.
- A relative clause is used approximately 20% of the time in post-modifiers.
 - More than 50% of the time, it is a *that* relative clause.
 - Nearly 20% of the time, the relativizer is omitted.
 - Sometimes, though not very often (5% of the time), it is a *which* relative clause.

people who work hard all their lives → noun + relative clause

facets of the university system → noun + *of* prepositional phrase

mistakes in life → noun + prepositional phrase

Sometimes more than one post-modifier can be used.

issues that are debated in the news → noun + *that* relative clause + prepositional phrase

Figure 7.13 Presentation of Noun Phrases with Both Pre- and Post-Modifiers

Noun phrases with both pre- and post-modifiers

A noun that has additional information is modified. The additional information is given with a modifier. When a noun has modifiers both before and after it, it has both pre- and post-modifiers.

the learning process in the university system is a noun phrase.

The noun is *process* and the additional information about the process is *the learning* and *in the university*. Because information is given both before and after the noun *process*, it has both a pre- and a post-modifier.

More examples of noun phrases with both pre- and post-modifiers:

different ways of learning

each element of the class

a thorough understanding of the subject

- More than 20% of nouns have both a pre- and a post-modifier.

the learning process in the university system →

the + adjective + **noun** + prepositional phrase

different ways of learning → adjective + **noun** + *of* prepositional phrase

each element of the class → adjective + **noun** + *of* prepositional phrase

a thorough understanding of the subject →

a + adjective + **noun** + *of* prepositional phrase

With this information available to teachers, the question arises as to what exactly they should do with it. The simple answer is to use it in class (in lieu of the noun clause chapter, perhaps). Introduce students to these patterns; have students identify the patterns in student writing models or their own work. Work on creating noun phrases as part of the editing process. This type of instructional material, however, is not always comfortable for a teacher who prefers language as absolute. But, perhaps ironically, when students encounter defined frequencies of use, they often feel more comfortable about such structures. Bennett and Bricker (2007) examined how learners and instructors felt, responded to, and valued currently unconventional corpus-based teaching and learning tools, such as the information presented in Figures 7.10-7.13; the findings suggest that students respond positively to this type of frequency instruction, appreciating the concrete aspects of language use that it provides. The frequencies allow a complex and chaotic abstract “monster” to feel routine and predictable. In addition, Garner (2011) found that student writing improved considerably after specific instruction in use and frequency with materials such as those above.

Given the data from Chapter 6, one thing is certain: the noun phrase will play a large role in the writing advanced level IEP students will encounter next: freshman composition. Pedagogical materials informed in part by the schematic representation presented in Chapter 6 will help students transition into their new discourse community.

7.5 Summary

In order to most efficiently assist students with their ascent up the staircase model of progression in academic writing skills, pedagogical materials (specifically published textbooks) serving EAP writers in the IEP should exploit certain methods and materials. Specifically, materials should focus on grammatical structures which characterize freshman

composition writing, as this writing level is the next that IEP students will encounter, evidenced in the staircase model of progression of academic writing skills. In addition, materials should provide models of student writing which can serve as “visible pedagogies” offering learners opportunities to engage with and grow into disciplinary possibilities. Furthermore, because learner writers are still developing proficiency in their L2, form is just as demanding as content, and the question is not whether or not to teach grammar but how best to do so. Frequency should play a key role in determining what grammar is taught, but frequency should also be considered in tandem with relevance. Basing the choice of what grammar to teach on the schematic representation presented in Chapter 6 allows materials writers and teachers to consider both. For example, as discussed above, even though appositives are used rather infrequently in academic writing, their inclusion in instructional materials is merited to provide students with explicit exposure to the structure; yet, instruction should also note the infrequency of the structure. Based on a review of the six most popular textbooks used in teaching IEP writing, current pedagogical materials address many of the relevant issues for this setting: the use of student models, a focus on both form and function, and attention to (mostly) relevant text types. Pedagogical materials do, however, need to provide more focus on phrasal structures, especially noun phrases.

CHAPTER EIGHT

CONCLUSION

8.1 Brief Overview

The design of this project was motivated by three concerns: the teaching of L2 writing in English for academic purposes (EAP), specifically in the intensive English program (IEP); the ubiquitous nature of freshman composition in North American higher education; and research on L2 writing in the English as a second language field, specifically corpus-based studies of academic writing. Given that instructional models allow potentially excluded participants an opportunity to become involved in academic discourse by facilitating participants' realization of the structures and functions exhibited in the targeted discourse and allowing for eventual independent construction of such discourse through experience and participation, I have argued that expert writing is not an appropriate model for novice writers. Purposes and audiences of expert and student writing differ; the expert standard is unfair and descriptively inadequate; and expert writing requires far transfer and thus is ineffective. Although multiple scholars have criticized student writing as inauthentic and showing a disregard for gatekeepers of discourse, I have shown that by informing EAP writing instruction and materials for the IEP on a staircase model of writing progression, student writing can be an effective model for learners; for EAP writers in the IEP, freshman composition as it is taught in North American higher education contexts is this aim. Methods and materials stemming from the proposed model are more likely to present writing that can be achievable for students, reflects their communicative purposes, and is also acceptable to the gatekeepers of the discourse, thereby allowing learners to engage with and grow into disciplinary possibilities and find their own academic voice as authors.

To confirm differences between the two levels of writing and serve as an informant to such materials, this study employed corpus-based methods to investigate the noun phrase in freshman composition and scholarly writing and reported differences between the two levels of writing in schematic representations. The representations highlighted the differences in frequencies of the noun phrase and its modifiers in freshman composition writing from that of scholarly writing as well as illustrated characteristic features of the noun phrase in successful freshman composition writing. It has also been argued that these features should be emphasized for EAP writers in the IEP who are seeking to begin undergraduate studies in North American contexts and suggestions for doing so have been provided.

8.2 Restatement of Research Aims

The interface of corpus linguistics and L2 writing has led to extensive corpus-based research focusing on the academic written register; however, the overwhelming majority of this corpus-based writing research has focused on the level of scholarly, or expert, writing. Because the purposes of expert and student writing differ, expert writing is not an appropriate model for novice writers. As discussed above, for EAP writers in the IEP with objectives to engage in undergraduate studies at institutions in North America, freshman composition is a more appropriate model of writing for four reasons:

1. all undergraduate students must confront this course, and therefore, it reflects authentic communicative purposes for IEP writers;
2. only one step higher on the staircase model of progression in academic writing skills (proposed in Chapter 2), freshman composition is the next level of writing IEP students will encounter;
3. the characteristic structures in freshman composition are a more attainable goal for IEP learners; and,

4. because freshman composition is itself a gatekeeper course, successful writing in that course is acceptable to gatekeepers of the discourse.

However, research into first year undergraduate writing is virtually non-existent. In the past decade, only approximately two corpus-based studies of writing have aimed to understand lexicogrammatical features which characterize first year undergraduate writing! Freshman composition must be researched so that instructors can be aware of the grammatical structures used in successful first year undergraduate writing so that they may ensure students understand it; descriptions are particularly important for learners to observe and understand how and why speakers use the language they do—not as a prescription for learner use—but as a presentation of data to enable learners to make their own informed choices. Because nouns and the noun phrase are integral structures in the written academic register, this study specifically compared the frequency of the noun phrase in freshman composition writing and scholarly writing with two main aims: to provide empirical evidence of the differences between the two levels of writing and to contribute to a description of first year undergraduate writing. Fulfilling both of these aims has, in turn, allowed positive contributions to more efficient EAP writing instruction in the IEP.

8.3 Review of Findings

From the compilation and comparison of profiles of scholarly writing and successful first year undergraduate writing, specifically in relation to the use of the noun phrase, the findings in this thesis showed that, indeed, distinct differences between the writing of successful freshman composition in first year undergraduate studies and that of scholarly writing in published books and articles written for an audience with a technical background do exist. These findings were illustrated in a schematic representation, displayed in Figures 8.1 and 8.2 below, for both levels of writing in order to facilitate comparisons between them.

Being able to “see” possible choices and the frequency with which those choices are realized can help students and instructors alike develop language skills.

Figure 8.1 Schematic Representation of the modified common head noun phrase in scholarly writing

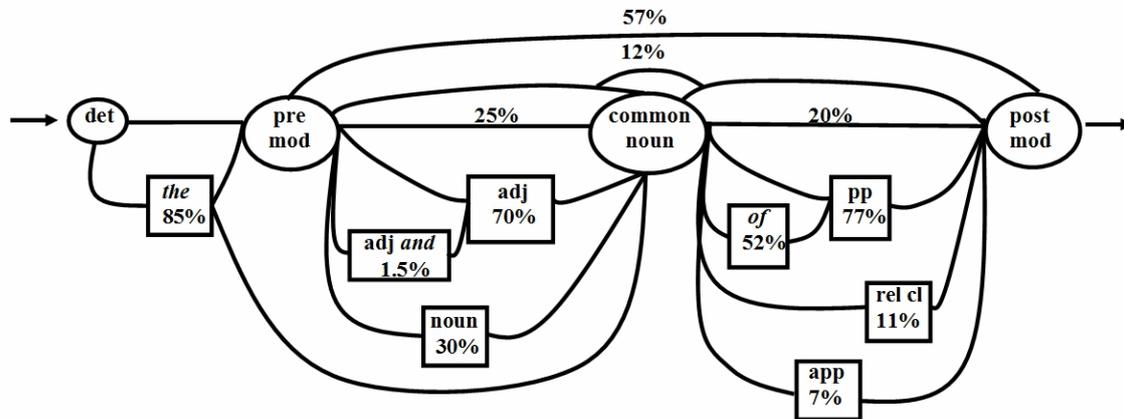
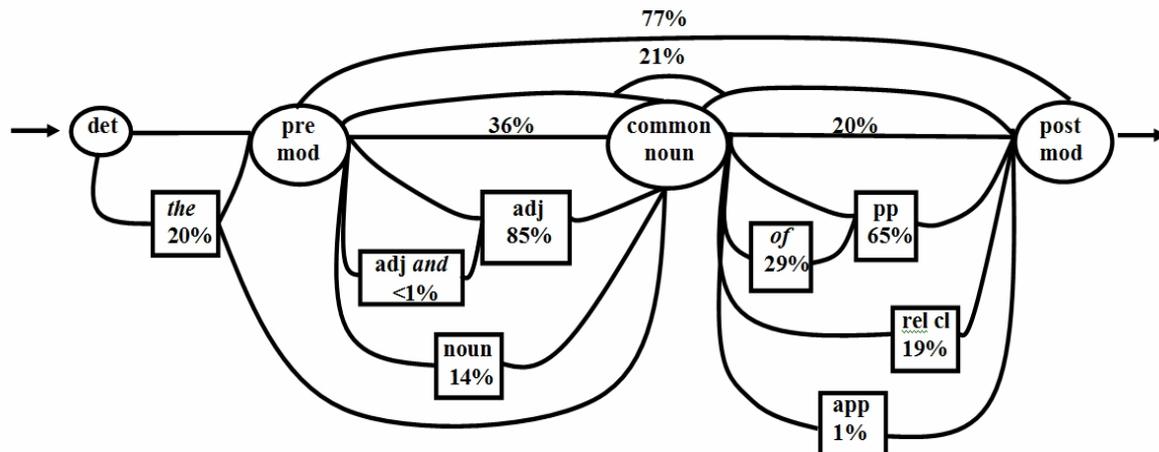


Figure 8.2 Schematic Representation of the modified common head noun phrase in the NAFWiC



Overall, general patterns of successful first year composition writing different from scholarly writing that emerged from the analysis include overall greater use of modifiers, less reliance on the definite article, greater use of common adjective pre-modifiers, and more diverse use of post-modifiers. Specifically, successful freshman composition writing shows a preference for modifying noun phrases, makes some use of relative clauses as post modifiers, frequently omits relativizers, prefers the use of *that* as a relativizer, relies heavily on

adjectives as pre modifiers, and makes no preference for use of *the* with head nouns. These results are important as they reveal distinct differences between scholarly and first year undergraduate writing that have heretofore gone unnoticed and demonstrate the complexity and variety of language use within and across disciplines and levels, confirming the arguments in this thesis that scholarly writing is a poor instructional model for EAP writers and that first year undergraduate writing merits its own description. Furthermore, a survey of current EAP writing textbooks in IEP settings revealed a need for the inclusion of such findings in order to maximize their effectiveness.

8.4 Significance of Findings

8.4.1 For corpus linguistics

These findings are significant for the field of corpus linguistics in two ways. Firstly, they provide a means of addressing one of the “plagues” of the progression of corpus linguistics in language teaching, that of applying corpus findings to classroom materials. Secondly, they highlight the need for more corpus-based studies to contribute to a description of first year undergraduate writing, furthering the research agenda of the discipline by exposing an area of language which is a valid target for studies investigating phraseology, frequency, and description in order to answer the two fundamental research questions the discipline broadly seeks to answer: What particular patterns are associated with lexical or grammatical features? How do these patterns differ within varieties and registers, or levels of writing proficiency?

Given the relevance of frequency in acquisition (Eillis, 2002), Biber and Reppen (2002) propose frequency should be considered in materials writing in order to determine what structures should be included in materials and to what extent, in what order structures should be presented, and what words should be included in those structures. Furthermore, for

corpus linguistics to revolutionize grammar teaching in the 21st century, Conrad (2000) recommended that frequency data alone not dictate pedagogy but be considered in tandem with relevance; that materials writers use corpus findings subtly; and, that teachers be made aware of the benefits of language description so that they are more willing to depart from a reliance on intuition and traditional grammar syllabi. By doing these things, she predicted that monolithic descriptions of English grammar would be replaced by register-specific descriptions; the teaching of grammar would become more integrated with the teaching of vocabulary; and, emphasis would shift from structural accuracy to the appropriate conditions of use for alternative grammatical constructions. Designing EAP writing materials based on data in schematic representations such as those in section 8.3 do these very things; specific suggestions for doing so were given in Chapter 7. Additional studies, such as those discussed in section 8.5 below, would continue this contribution.

Lexicogrammatical structures do differ across register; they differ across disciplines; they differ across genres—we see clearly now that they also differ across levels. And not just between L2 writers and L1 writers, but between novice and expert writing regardless of L1.

8.4.2 For L2 writing

These findings are also significant for the field of L2 writing in two ways. They provide a reliable answer to the “what should we teach” debate for EAP writing instruction in the IEP and demonstrate that student writing is a valid genre in the discipline of freshman composition that should be used (more so than expert writing) to inform writing instruction.

Hyland (2004), among others, argues against the “common-core” teaching approach to EAP writing given the significant differences in text types and lexicogrammatical structures in disciplines; in fact, the findings discussed in Chapter 3 section 3.5 confirm this. Spack (1998) also questions the “common-core” teaching approach to EAP writing, arguing that L2

writing researchers and teachers should leave the teaching of writing in the disciplines to the instructors of those disciplines, as the goal of an EAP writing program should be to prepare students to become better academic writers in general. As scholars across the decades agree, however, determining what academic writing is and what EAP writers need to know in order to produce it is no easy task (Johns, 1988; Spack, 1998; Coxhead & Byrd, 2007; Hinkel, 2011). Yet, particular assignments are common and recurring in freshman composition, and over time they have become a genre, which uses lexicogrammatical features differently than other genres; therefore, freshman composition is defined as a discipline based on its courses, assignments, and distinctive realization of shared features in them. Given that freshman composition is the next step in the staircase of progression in academic writing skills, it is, indeed, possible to determine what EAP writers need to know in order to produce successful writing in the next discipline they will encounter. This study specifically highlights the following features as among the first which should be taught to EAP writers:

- the noun phrase and four types thereof;
- nouns with only a determiner, nouns with a pre-modifier only, nouns with a post-modifier only, and nouns with both a pre-modifier and a post-modifier;
- for nouns with a determiner only, all available choices should be highlighted (for the definite noun phrase, a definite determiner, a possessive determiner, or a demonstrative determiner);
- for nouns with a pre-modifier only, all available pre-modifier types should be presented with a focus on adjectives as pre-modifiers;
- for nouns with a post-modifier only, all available post-modifier types should be presented, but instruction should focus on the use of *of*- and other prepositional

phrases as well as relative clauses using a *that* relativizer or omitting the relativizer;

- awareness that noun phrases may have both a pre-modifier and a post-modifier.

Teacher training materials as well as student classroom materials should highlight these structures.

Although the proposed model and research in this thesis is directly applicable to EAP students studying in an IEP with aspirations to undertake North American undergraduate studies, every academic context is different. In some countries—in fact, probably most outside of North America—there are no freshman composition programs, and, therefore, the significance of the research discussed here may not be directly applicable in these contexts. Nevertheless, the proposed staircase model holds true for any EAP context: students who are studying in pre-sessional or other courses at level 1 on the staircase with aims to engage in undergraduate study at level 2 should engage in materials based on the student writing genre, which will help propel them to step 2, not step 4 or even 3.

8.5 Limitations of Study

Two limitations of the study should be considered. Firstly, though the use of corpus-cited references as a data source negated the need for creation of a scholarly writing corpus specifically for this study (saving time and resources) and allowed for more data than the author would have been able to gather independently, this method limited the study in two ways. For one, the comparisons made between the two types of writing were limited to that data which was published (e.g. distribution of lexical word class) in the corpus-cited references. As such, comparisons of other types (e.g. sentence length) could not be made because no such data for scholarly writing was published. Similarly, because data in the corpus-cited texts is generally provided in chart and graph form, precise percentages and

figures often can not be ascertained. Thus, the reader has to make a guess as to the approximate frequencies for certain features (see Figures 6.4 and 6.5 in section 6.3, for example). Although the graphs are clear and very close approximations can be made, and in this study the differences between the data were substantial enough that the results would not have changed if exact figures were obtained, this limitation must be acknowledged, particularly for future research where such precision may make a significant difference.

An additional limitation of the study is that no statistical tests were used to establish significance in the differences between freshman composition writing and scholarly writing. As Tribble (2002) argues (and noted in section 6.4.3), percentages provide a useful rough means of differentiating between texts. The percentages reported in Chapter 6 do confirm a difference in the NAFWiC and scholarly writing, but tests of statistical significance could be used to confirm the differences and provide further evidence for the argument presented here.

8.6 Further Research

Given the individual audiences, topics, and education of the authors, it is no surprise that the use of the noun phrase in successful freshman composition writing differs markedly from scholarly writing. With texts of scholarly writing written by authors with a graduate or postgraduate education to readers of a technical audience, the language choices, perhaps, should differ from the choices of first year undergraduates who are writing for an instructor. The findings and significance of those findings necessitate that more research be undertaken of successful freshman composition writing, both in terms of more texts and more features and functions, to better prepare EAP instructors to help their students enter the undergraduate discourse community. Although this study contributed to a description of freshman composition, we are a long way from a complete understanding of the differences between expert and novice writing and an understanding of what structures characterize the latter, and

therefore, should be highlighted in EAP writing materials in the IEP. The differences of use in the noun phrase in these two levels of writing signal the need for more research on other aspects of successful freshman composition, e.g. the verb phrase, role of adjectives and adverbs, lexical bundles, and even more research on the noun phrase for understanding the connection to topic and/or style, to form a more complete representation of the genre/discipline.

Investigating additional features and functions of successful first year undergraduate writing (such as those discussed above), expansion of the NAFWiC and examination of more texts and associated features and functions in each of the texts would be ideal. Furthermore, different samples of writing may provide additional insights into distinct linguistic features and functions of second year undergraduate writing, for example, or American versus British undergraduate writing. As one of the first studies conducted on successful first year undergraduate writing, this investigation has revealed notable information pertaining to such writing, but must be ensued with further research to provide a more complete representation of successful freshman composition writing which can authoritatively inform L2 writing materials development and teacher training.

Several productive studies involving a range of approaches lend themselves to the NAFWiC for further research. The first and perhaps most obvious studies for future research are those resulting from the findings in Chapter 6, especially those which highlight differences between the two types of writing. These include 1) Do pre-modifier types increase (proportionally) with level? 2) What functions does the *of*-phrase realize in the NAFWiC? How do these compare to those in scholarly writing? 3) What functions are noun phrases performing in the NAFWiC? Is there an alternate function that could account for

their increased frequency in freshman composition versus scholarly writing? 4) What determiners are being used in the NAFWiC and what functions do they complete?

Any number of other studies naturally follow the research presented here, including further investigation of lexicogrammatical features in freshman comp; for example, as mentioned above, examining the verb phrase or learning more about the role of adjectives and prepositions in the NAFWiC. A key-word analysis could also prove an interesting corpus-based follow up study, shedding light on what freshman composition is “about” (Scott, 2000). Using the MICUSP as a reference corpus could potentially identify differences between step 2a and step 2b and/or step 3 writing on the staircase model. Following this, corpus-based studies which have investigated lexicogrammatical features of specific disciplines could be repeated in the NAFWiC to provide further evidence of freshman comp as its own discipline. While corpora the size of the NAFWiC are generally sufficient for these types of studies, as discussed in Chapter 4, growing the NAFWiC could also prove a productive endeavor for future research involving other approaches. Adding contributions from additional freshman composition courses would serve to confirm or provide new details on the text types and practices embodied in the courses. Furthermore, a genre-based study, one key approach to L2 writing research as discussed in Chapter 3, could reveal additional details about freshman composition to inform pedagogical practices and materials. Finally, the results from corpus-based studies involving the NAFWiC may be used in conjunction with results from corpus-based studies on learner corpora to create the most effective and efficient pedagogical practices and materials. Gilquin et al (2007) and Hinkel (2003), among others, advocate the greater use of learner corpora to inform L2 writing as these studies reveal specific errors made by learners. Understanding the typical inter-language of a specific group of learners (learner corpora) combined with the typical lexicogrammatical features of the discipline these learners

will be entering (NAFWiC) could provide the most effective and efficient methods and materials possible for teaching grammar for EAP writing in the IEP.

This study has highlighted the need for more research into lexicogrammatical features of first year undergraduate writing and any number of studies could contribute to this new corpus-based description of academic writing.

REFERENCES

- AAIEP. (n.d.). American Association of Intensive English Programs. [online] Available at <<http://www.aaiep.org/>> [Accessed 9 July 2010].
- Aarts, J. (2002). Does corpus linguistics exist? Some old and new issues. In L. E. Breivik & A. Hasselgren (Eds.), *From the COLT's mouth ... and others': Language corpora studies in honour of Anna-Brita Stenström*, p. 1-19. Amsterdam: Rodopi.
- Ädel, A. & Garretson, G. (2006). Citation practices across the disciplines: The case of proficient student writing. In M.C. Pérez-Llantada Auría, R. Pló Alastrué, & C. P. Neumann, *Academic and professional communication in the 21st century: Genres, rhetoric and the construction of disciplinary knowledge. Proceedings of the 5th International AELFE Conference.* Zaragoza: Universidad de Zaragoza.
- Ädel, A. & Reppen, R. (2008). *Corpora and discourse: The challenges of different settings*. Amsterdam: John Benjamins.
- Adolphs, S. (2006). *Introducing electronic text analysis: A practice guide for language and literacy studies*. London, UK: Routledge.
- Ahrenhoerster, G. (2006). Will they still respect us in the morning? A study of how students write after they leave the composition classroom. *Teaching English in the Two Year College*, 34(1), p. 20-31.
- Aktas, R. N., & Cortes, V. (2008). Shell nouns as cohesive devices in published and ESL student writing. *Journal of English for Academic Purposes*, 7(1), p. 3-14.
- Albers, S. (2007). The VIKING corpus of academic student writing: Documentation release 1. Portland, OR: Portland State University.

- Ali, Z. & Yunus, M. (2004). An ESL writing course: Unravelling students' needs and concerns. *MELTA: The English Teacher XXXIII*.
<http://www.melta.org/my/ET/2004/2004-114.pdf>
- Allison, L, Bryant, L, & Hourigan, M. (Eds.). (1997). *Grading in the post-process classroom: From theory to practice*. Portsmouth, NH: Boynton/Cook.
- Apple, M. W. (1995). *Education and power*. London, UK: Routledge.
- Archibald, A. & Jeffrey, G. C. (2000). Editorial: Second language acquisition and writing: A multi-disciplinary approach. *Learning and Instruction, 10*, p. 1-11.
- Atkinson, D. (2000). On Robert B. Kaplan's response to Terry Santos et al.'s "On the future of second language writing." *Journal of Second Language Writing, 9*(3), 317-320.
- Azar, B. S. (2002). *Understanding and using English grammar*. 3rd ed. New York: Pearson Education.
- Bacha, N. N. (2010). Teaching the academic argument in a university EFL environment. *Journal for English for Academic Purposes, 9*(3), p. 229-241.
- Baker, B. (2008). L2 Writing and L1 Composition in English: Towards an Alignment of Effort. *McGill Journal of Education, 43*(2), p. 139-156.
- Baker, P. (2010). *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press.
- Baker, P. (2006). *Using corpora in discourse analysis*. London, UK: Continuum.
- Banerjee, S. (2008). Revisiting argumentation: A comparison of successful and unsuccessful NES and ESL college writing. Unpublished dissertation. ProQuest document ID: 3300328.
- Baratta, A. (2009). Revealing stance through passive voice. *Journal of Pragmatics, 41*, p. 1406–1421.

- Barlow, M. (2011). Corpus linguistics and theoretical linguistics. *International Journal of Corpus Linguistics*, 16(1), p. 3–44
- Bartlett, T. (2003). Why Johnny can't write, Even though he went to Princeton. *Chronicle of Higher Education*, 49(17), pA39-A40.
- Bazerman, C. (2006). Distanced and refined selves: Educational tensions in writing with the power of knowledge. In M. Hewings, (Ed.), *Academic writing in context*, p. 23-29. Continuum: London.
- Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. London: Longman.
- Bhatia, V. K. & Candlin, C. N. (Eds.). (2001). *Teaching English to meet the needs of business education in Hong Kong*. Hong Kong: Department of English and Communication, City University of Hong Kong.
- Beaufort, A. (1999). *Writing in the real world: Making the transition from school to work*. New York, NY: Columbia University Teachers College Press.
- Belcher, D. & Braine, G. (Eds.). (1995). *Academic writing in a second language*. Norwood, NJ: Ablex.
- Belles-Fortuño, B. & Campoy-Cubillo, M.C. (2010). 'I sort of feel like, um, I want to, agree with that for the most part': Reporting intuitions and ideas in spoken academic discourse. In M. C. Campoy-Cubillo, B. Belles-Fortuño & L. Gea-Valor (Eds.), *Corpus-based Approaches to English Language Teaching*, p. 56-66. London, UK: Continuum.
- Benesch, S. (2001). *Critical English for academic purposes: Theory, politics, and practice*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bennett, G. (2010). *Using corpora in the language learning classroom*. Ann Arbor, MI: University of Michigan Press.

- Bennett, G. & Bricker, M. (2007). An Affective Look at Corpus-Driven Activities. Presentation at the 41st Annual Teaching English to Speakers of Other Languages (TESOL) Convention, Seattle, Washington, 23 March, 2011
- Bennett, G., Frodesen, J., Schmitt, D., Keck, K., & Wald, M. (2010). Re-imagining Pedagogy and Practice for Writing from Sources. Presentation at the 44th Annual Teaching English to Speakers of Other Languages (TESOL) Convention, Boston, Massachusetts, 25 March, 2010.
- Bennett, G., Frodesen, J., Schmitt, D., & Wald, M. (2011). Activities for Putting the “E” Back in EAP Writing. Presentation at the 45th Annual Teaching English to Speakers of Other Languages (TESOL) Convention, New Orleans, Louisiana, 18 March, 2011.
- Bernhardt, E. B., Edwards, P. R., & Wojahn, P. G. (1989). Teaching college composition with computers: A program evaluation study. *Written Communication*, 6, p. 108-133.
- Bernhardt, E. B., Wojahn, P. G., & Edwards, P. R. (1990). Teaching college composition with computers: A timed observation study. *Written Communication*, 7, p. 342-374.
- Berkenkotter, C. & Huckin, T. (1995). *Genre knowledge in disciplinary communication: Cognition/Culture/Power*. Hillsdale, NJ: Erlbaum.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge, UK: Cambridge University Press.
- Biber, D. (1995). *Dimensions of register variation*. Cambridge, UK: Cambridge University Press.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins Publishing.
- Biber, D. (2009). A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistics*, 14(3), p. 275–311.

- Biber, D. & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), p. 263-286.
- Biber, D. & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson*, p. 181–190. Amsterdam: Rodopi.
- Biber, D. & Conrad, S. (2001). Quantitative corpus-based research: Much more than bean-counting. *TESOL Quarterly*, 35(2), p. 331-336.
- Biber, D. & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, 24, p. 199-208.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: lexical bundles in university teaching and textbooks. *Applied Linguistics* 25(3), p. 371-405.
- Biber, D. Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45(1), p. 5-35.
- Biber, D., Csomay, E., Jones, J., & Keck, C. (2004). A corpus linguistic investigation of vocabulary-based discourse units in university registers. In U. Connor & T. A. Upton (Eds.), *Applied corpus linguistics: A multi-dimensional perspective*, p. 53-72. Amsterdam: Rodopi.
- Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *TESOL Quarterly* 35(1), p. 9-48.
- Biber, D., Connor, U., & Upton, T. (2007). *Discourse on the move*. Amsterdam: John Benjamins.

- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., (1999). *Longman grammar of spoken and written English*. Pearson Education: New York.
- Bitchener, J. & Knoch, U. (2010). Raising the linguistic accuracy of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, 19(4), p. 202-217.
- Bizzell, P. (1982). College Composition: Initiation into the Academic Discourse Community. *Curriculum Inquiry*, 12(2), p. 191-207.
- Bloor, M. & Bloor, T. (2006). There'll be some changes made: Predicting future events in academic and business genres. In M. Hewings, *Academic writing in context*, p. 182-198. Continuum: London.
- Bondi, M. & Scott, M. (2010). *Keyness in texts*. Amsterdam: John Benjamins.
- Bowker, L. & Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora*. London, UK: Routledge.
- Braun, S., Kohn, K., & Mukherjee, J. (2006). *Corpus technology and language pedagogy*. Frankfurt: Peter Lang.
- Brezinsky, J. (2010). *Writing Textbooks*. [email] (Personal communication, 6 October 2010).
- The British Council. (1975). English for Academic Study, With Special References to Science and Technology: Problems and perspectives. An ETIC Occasional Paper. London: English Teaching Information Centre.
- Brown, H. D. (2001). *Teaching by principles*. New York, NY: Longman.
- Brown, H. D. (2007). *Principles of language learning and teaching*. White Plains, NY: Pearson Education.
- Brown, J.D. (1995). *The elements of language curriculum*. Boston, MA: Heinle & Heinle.

- Brown University. (2010). *About Brown*. [online] Available at
 <<http://www.brown.edu/web/about/history>> [Accessed 9 July 2010].
- Bruce, I. (2010). Textual and discursal resources used in the essay genre in sociology and English. *Journal of English for Academic Purposes*, 9(3), p. 153-166.
- Bussmann, H., Kazzazi, K., & Trauth, G. (1998). *Routledge dictionary of language and linguistics*. London: Routledge.
- Bulter, C. (1990). Qualifications in science: Modal meanings in scientific text. In W. Nash (Ed.), *The Writing Scholar*, p. 137-170. London, UK: Sage Publications.
- Butler, C., Gómez-González, M., Doval Suárez, S.M. (2005). *The dynamics of language use*. Amsterdam: Rodopi.
- Butvilofsky, S. A. (2010). Towards the development of a biliterate pedagogy: A case study of emerging biliterate writing. *ProQuest document ID: 755473255*.
- Byrd, H.P. (1994). Writing grammar textbooks: Theory and practice. *System*, 22(1), p. 245-255.
- Cambridge University Press. (2010). *Writers at Work—English as a second language*. [online] Available at
 <http://www.cambridge.org/us/esl/catalog/subject/project/item5600130/writers-at-work/?site_locale=en_US> [Accessed 5 December 2010].
- Cambridge University Press. (2011). *Cambridge International Corpus*. [online] Available at
 <http://www.cambridge.org/us/esl/catalog/subject/custom/item3637700/Cambridge-International-Corpus-Cambridge-International-Corpus/?site_locale=en_US> [Accessed 15 February 2011].
- Campoy, M. C., Belles-Fortuno, B., & Gea-Valor, M. L. (2010). *Corpus-based approaches to English language teaching*. London, UK: Continuum.

- Canagarajah, S. (2002). *Critical academic writing*. Ann Arbor: University of Michigan Press.
- The Carnegie Committee. (Anon., 1899). [online] *Cornell Alumni News*, II(10). Available at <http://ecommons.library.cornell.edu/bitstream/1813/3164/11/002_10.pdf> [Accessed on 9 July 2010].
- Carroll, L. (2002). *Rehearsing new roles: How college students develop as writers*. Carbondale, IL: Southern Illinois University Press.
- Carter, D. (1993). Critical thinking for writers: Transferable skills or discipline-specific strategies? *Composition Studies/Freshman English News*, 21(1), p. 86-93.
- Carter, R. & McCarthy, M. (2006). *Cambridge grammar of English: A comprehensive guide to spoken and written English grammar and usage*. Cambridge: Cambridge University Press.
- Casanave, C. P. (2004). *Controversies in L2 writing: Dilemmas and decisions in research and instruction*. Ann Arbor, MI: University of Michigan Press.
- Casanave, C. P. (2010). Taking risks? A case study of three doctoral students writing qualitative dissertations at an American university in Japan. *Journal of Second Language Writing*, 19(1), p. 1-6.
- Charles, M. (2003). 'This mystery...': A corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes*, 2, p. 313-326.
- Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes* (25)3, p. 310-331.

- Charles, M. (2007). Argument or evidence? Disciplinary variation in the use of the noun that pattern in stance construction. *Journal of English for Academic Purposes*, 26, p. 203-218.
- Charles, M. (2009). Stance, Interaction and the Rhetorical Patterns of Restrictive Adverbs: Discourse Roles of *Only*, *Just*, *Simply* and *Merely*. In Charles, M., Pecorari, D., & Hunston, S. (Eds.), *Academic writing: At the interface of corpus and discourse*, p. 152-169. London: Continuum.
- Charles, M. (2011). Adverbials of result: Phraseology and functions in the Problem-Solution pattern. *Journal of English for Academic Purposes*, 10, p. 47-60.
- Charles, M., Pecorari, D., & Hunston, S. (Eds.). (2009). *Academic writing: At the interface of corpus and discourse*. London: Continuum.
- Chen, Y. H. & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), p. 30-49.
- Cheng, W., Warren, M. & Xun-feng, X. (2003). The language learner as language researcher: Putting corpus linguistics on the timetable. *System*, 31, p. 173-186.
- Cho, S. (2010). Academic biliteracy challenges: Korean scholars in the United States. *Journal of Second Language Writing*, 19(2), p. 82-94.
- Clanchy, R. & Ballard, B. (1992). *How to write essays*. Melbourne: Longman Chesire.
- Coniam, D. (2004). Concordancing yourself: A personal exploration of academic writing. *Language Awareness*, 13(1), P. 49-55.
- Conrad, S. (1999). The importance of corpus-based research for language teachers. *System*, 27, p. 1-18.
- Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly*, 34(3), p. 548-560.

- Conrad, S. & Biber, D. (2009). *Real grammar*. White Plains, NY: Pearson Longman.
- Conroy, M. A. (2010). Internet tools for language learning: Students taking control of their writing. *Australian Journal of Educational Technology*, 26(2), p. 861-882.
- Cooper, C. & Odell, L. (Eds.). (1998). *Evaluating writing: The role of teachers' knowledge about text, learning, and culture*. Urbana, IL: National Council of Teachers of English.
- Cortes, V. (2002). Lexical bundles in freshman composition. In R. Reppen, S. M. Fitzmaurice, & D. Biber (Eds.), *Using corpora to explore linguistic variation*, p. 131-145. Amsterdam: John Benjamins.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), p. 397-423.
- Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, 17(4), p. 391-406.
- Cortes, V. & Csomay, E. (2007). Positioning lexical bundles in university lectures. In M. C. Campoy & M. J. Luzon (Eds.), *Spoken corpora in applied linguistics*, p. 57-76. Bern: Peter Lang.
- Cortes, V. & Csomay, E. (2009). Lexical bundle distribution in university classroom talk. In S. T. Gries, S. Wulff, & M. Davies (Eds.), *Corpus-linguistic applications: Current studies, new directions*, p. 153-168. Amsterdam: Rodopi.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, p. 213-238.
- Coxhead, A. & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing* (16), p. 129-147.
- Deleger, L. Merkel, M., & Zweigenbaum, P. (2009). Translating medical terminologies through word alignment in parallel text corpora. *Journal of Biomedical Informatics*, 42(4), p. 692-701.

- Detterman, D. (1993). The case for the prosecution: Transfer as an epiphenomenon. In D. Detterman & R. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction*, p. 1-24. Norwood, NJ: Ablex.
- Diani, G. (2008). Emphasizers in spoken and written academic discourse. *International Journal of Corpus Linguistics*, 13(3), p. 296–321.
- di Gennaro, K. (2006). Second language writing ability: Towards a complete construct definition. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, 6(2), p. 1-17.
- Dikli, S. (2010). The nature of automated essay scoring feedback. *CALICO Journal*, 28(1), p. 99-134.
- Diller, C. & Oates, C. (2002). Infusing disciplinary rhetoric into liberal education: A cautionary tale. *Rhetoric Review*, 21, p. 53-61.
- Driscoll, C. (2010). *Academic Writing*. [email] (Personal communication, 27 October 2010).
- Downs, D. & Wardle, E. (2007). Teaching about Writing, Righting Misconceptions: (Re)envisioning "First-Year Composition" as "Introduction to Writing Studies." *College Composition and Communication*, 58(4), p. 552-584.
- Dudley-Evans, T. (1993). Variation in communication patterns between discourse communities: The case of highway engineering and plant biology. In G.M. Blue (Ed.), *Language, Learning and Success*, p. 141-147. Hemel Hempstead: Phoenix ELT.
- Duguid, A. (2009). Loud signatures: Comparing evaluative discourse styles–patterns in rants and riffs. In U. Römer & R. Schulze (Eds.), *Exploring the lexis–grammar interface*, p. 289-315. Amsterdam: John Benjamins.
- Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, 28(3), p. 157–169.

- Durrant, P. & Mathews-Aydmh, J. (2011). A function-first approach to identifying formulaic language in academic writing. *English for Specific Purposes*, 30(1), p. 58-72.
- Durst, R. K. (1999). *Collision course: Conflict, negotiation, and learning in college composition*. Urbana, IL: National Council of Teachers of English.
- Elbow, P. (1998). Reflection on academic discourse: How it relates to freshmen and colleagues. In V. Zamel & R. Spack (Eds.), *Negotiating academic literacies*, p. 145-171. Mahwah, NJ: Lawrence Erlbaum.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, p. 143-188.
- Ellis, N. C., & Larsen-Freeman, D. (2009). Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage. *Language Learning*, 59(1), p. 93-128.
- ElMalik, A. & Nesi, H. (2008). Publishing research in a second language: The case of Sudanese contributors to international medical journals. *Journal of English for Academic Purposes* 7, p. 87-96.
- Enright, M. K. & Quinlan, T. (2010). Complementing human judgment of essays written by English language learners with e-rater scoring. *Language Testing*, 27(3), p. 317-334.
- Evans, S. & Green, C. (2007). Why EAP is necessary: A survey of Hong Kong tertiary students. *Journal of English for Academic Purposes*, 6, p. 3-17.
- Fairclough, N. (2000). *New Labour, new language*. London, UK: Routledge.
- Feak, C., Reinhart, S., & Rohlck, T. (2009). *Academic interactions*. Ann Arbor, MI: University of Michigan Press.
- Ferris, D. R. (1994). Rhetorical strategies in student persuasive writing: Differences between

- native and non-native English speakers. *Research in the Teaching of English*, 28(1), p. 45-65.
- Ferris, D. R. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, 8(1), p. 1-12.
- Ferris, D. R. (2004). The “Grammar Correction” Debate in L2 Writing: Where are we, and where do we go from here? (and what do we do in the meantime...?) *Journal of Second Language Writing*, 13, p. 1-14.
- Ferris, D. R. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. *Studies in Second Language Acquisition*, 32(2), p. 181-201.
- Ferris, D. R. (2011). Expectations & Challenges for L2 Students in Undergraduate Writing Programs. Presentation at the 45th Annual Teaching English to Speakers of Other Languages (TESOL) Convention, New Orleans, Louisiana, 18 March, 2011.
- Ferris, D. R. & Hedgcock, J. (2005). *Teaching ESL composition: Purpose, process, and practice*. Mahwah, NJ: Lawrence Erlbaum.
- Ferris, D. R., Brown, J. C., Liu, H., & Stine, M. E. A. (in press). Responding to L2 students in college writing classes: What teachers say and what they do. *TESOL Quarterly*.
- Flowerdew, J. (1993). An educational, or process, approach to the teaching of professional genres. *ELT Journal*, 47, p. 305-316.
- Flowerdew, J. (1996). Concordancing in language learning. In M.C. Pennington (Ed.), *The power of CALL*, p. 97-113. Houston, TX: Athelstan.
- Flowerdew, J (Ed.). (2002). *Academic discourse*. Harlow, England: Pearson Education.
- Flowerdew, J. & Forest, R. W. (2009). Schematic structure and lexico-grammatical realization in corpus-based genre analysis: The case of *research* in the PhD literature

- review. In Charles, M., Pecorari, D., & Hunston, S. (Eds.), *Academic writing: At the interface of corpus and discourse*, p. 15-36. London: Continuum International Publishing Group.
- Flowerdew, J. & Peacock, M. (2001). *Research perspectives on English for academic purposes*. Cambridge, UK: Cambridge University Press
- Flowerdew, L. (2002). Corpus-based analysis in EAP. In J. Flowerdew (Ed.), *Academic discourse*, p. 95-114. Harlow, England: Pearson Education.
- Flowerdew, L. (2008). *Corpus-based analysis of the problem-solution pattern*. Amsterdam: John Benjamins.
- Folse, K. & Pugh, T. (2010). *Greater essays*. Boston, MA: Heinle Cengage Learning.
- Folse, K. & Muchmore-Vokoun, A., & Vestri Solomon, E. (2010). *Great essays*. Boston, MA: Heinle Cengage Learning.
- Francis, B., Robson, J., & Read, B. (2010). An analysis of undergraduate writing styles in the context of gender and achievement. *Studies in Higher Education*, 26(3), p. 313-326.
- Frankenberg-Garcia, A. (2006). Raising teachers' awareness of corpora. *Language Teaching*, Available on CJO.
- Frawley, W. (1997). *Vygotsky and cognitive science: Language and the unification of the social and computational mind*. Cambridge, MA: Harvard University Press.
- Freedman, A. (1995). The What, Where, When, Why, and How of Classroom Genres. In J. Petraglia (Ed.), *Reconceiving writing rethinking writing instruction*, p. 121-144. Mahwah, NJ: Laurence Erlbaum.
- Freire, P. (1972). *Pedagogy of the oppressed*. London, UK: Penguin.
- Friginal, E. (2009). *The language of outsourced call centers: A corpus-based study of cross-cultural interaction*. Amsterdam: John Benjamins.

- Friginal, E. (2010). *Discussion on our research involving corpus methods*. [email] (Personal communication, 17 March 2010).
- Frodesen, J. (.2001). Grammar in Writing. In M. Celce-Murcia, (ed.), *Teaching English as a second or foreign language*, 3rd edition, p. 233-248. Boston, MA: Heinle & Heinle.
- Frodesen, J. & Holton, C. (2003). Grammar and the ESL writing class. In B. Kroll (Ed.), *Exploring the dynamics of second language writing*, p. 141-161. New York, NY: Cambridge University Press.
- Fulkerson, R. (2005). Composition at the Turn of the Twenty-First Century. *College Composition and Communication*, 56(4), p. 654-687.
- Gabrielatos, C. (2005). Corpora and language teaching. Just a fling or wedding bells? *TESL-EJ* 8(40), A1.
- Galloway, I. (2005). Computer learner corpora and their pedagogical application. *TESOL Quarterly*, 39(2), p. 333-339.
- Gage, J. T. (1982). Freshman English: In Whose Service? *College English*, 44(5), p. 469-474.
- Gardner, S. (2008). Mapping ideational meaning in a corpus of student writing. In C. Jones & E. Ventola, (Eds), *New developments in the study of ideational meaning: From language to multimodality*, p. 169-188. London, UK: Equinox Publishing.
- Gardner, S. (2009). Extended abstract: Evaluation across disciplinary groups in university student writing: The critique genre family as texts and text. In: Edwardes, M. (Ed.), *Proceedings of the BAAL Annual Conference 2008. Taking the Measure of Applied Linguistics, Swansea University*, p. 47-50. London: BAAL/ Scitsiugnill Press.
- Gardezi, S. A. & Nesi, H. (2009). Variation in the writing of economics students in Britian and Pakistan: The case of conjunctive ties. In M. Charles, D. Pecorari, & S. Hunston

- (Eds.), *Academic writing: At the interface of corpus and discourse*, p. 236-250.
London: Continuum International Publishing Group.
- Garner, J. (2011). *Dissertation results from adverbials in writing*. [email] (Personal communication, 25 March 2011).
- Gaskell, D. & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32, p. 301–319.
- Gavioli, L. (2005). *Exploring corpora for ESP teaching*. Philadelphia: John Benjamins.
- Gavioli, L. (2001). The learner as researcher: Introducing corpus concordancing in the classroom. In G. Aston, (Ed.), *Learning with corpora*, p. 108-137. Houston, TX: Athelstan.
- Ghadessy, M., Henry, A., & Roseberry, R. (Eds.). (2001). *Small corpus studies and ELT: Theory and practice*. Amsterdam: John Benjamins.
- Giesler, C. (1994). *Academic literacy and the nature of expertise: Reading, writing, and knowing in academic philosophy*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gillett, A. (1996). What is EAP? *IATEFL ESP SIG Newsletter*, 6, 17-23.
- Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, 6, p. 319–335.
- Ginther, A. & Grant, L. (1996). *A Review of the Academic Needs of Native English-Speaking College Students in the United States*. Princeton, New Jersey: Education Testing Service.
- Giroux, H. A. (1981). *Ideology, culture and the process of schooling*. London, UK: Falmer Press.
- Gledhill, C. (2000). The discourse function of collocation in research article introductions. *English for Specific Purposes*, 19(2), p. 115-135.

- Gledhill, C. (1996). Science as a collocation: Phraseology in cancer research articles. In S. Botley, et al, (Eds.), *Proceedings of Teaching and Language Corpora Conference 1996 UCREL 9*, p. 108-126. UK: University of Lancaster.
- Goldberg, A. E. (2006). *Constructions at work. The nature of generalization in language*. Oxford, UK: Oxford University Press.
- Golebiowski, Z. (2009). Prominent messages in Education and Applied Linguistic abstracts: How do authors appeal to their prospective readers? *Journal of Pragmatics*, 41, p. 753–769.
- Gómez González, M., Lachlan Mackenzie, J., & González Álvarez, E. M. (2008). *Languages and cultures in contrast and comparison*. Amsterdam: John Benjamins.
- Grabe, W. & Kaplan, R. (1996). *Theory and practice of writing*. London, UK: Longman.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. Cowie (Ed.), *Phraseology: Theory, analysis and applications*, p. 145-160. Oxford, UK: Oxford University Press.
- Granger, S. & Paquot, M. (2009). In search of a general academic English: A corpus-driven study. In K. Katsampoxaki-Hodgetts (Ed.), *Options and practices of LSP practitioners, University of Crete, 7-8 February 2009*, p. 94-108. Crete: University of Crete Publications.
- Granger, S. & Rayson, P. (1998). Automatic profiling of learner texts. In S. Granger (Ed.), *Learner English on computer*, p. 119-131. London, UK: Longman.
- Granger, S., Lerot, J., & Petch-Tyson, S. (2003). *Corpus-based approaches to contrastive linguistics and translation studies*. Amsterdam: Rodopi.
- Graves, K. (2000). *Designing language courses: A guide for teachers*. Boston, MA: Heinle & Heinle

- Gray, B. (2010). On the use of demonstrative pronouns and determiners as cohesive devices: A focus on sentence-initial this/these in academic prose. *Journal of English for Academic Purposes*, 9, p. 167-183.
- Gray, B. & Cortes, V. (2011). Perception vs. evidence: An analysis of this and these in academic prose. *English for Specific Purposes*, 30, p. 31–43.
- Gregg, K. (1984). Krashen’s monitor and Occam’s razor. *Applied Linguistics*, 5, p. 79-100.
- Greenbaum, S. (1974). Some verb-intensifier collocation in American and British English. *American Speech*, 49, p. 79-89.
- Gries, S. (2010). Corpus linguistics and theoretical linguistics: A love–hate relationship? Not necessarily... *International Journal of Corpus Linguistics* 15(3), p. 327–343.
- Groom, N. (2005). Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes*, 4(3), p. 257–277.
- Hadley, G. (2002). Sensing winds of change: An introduction to data-driven learning. *RELC Journal*, 33(2), p. 99-124.
- Hale, G., Taylor, C., Bridgeman, B., Carson, J., Kroll, B. & Kanto, R. (1996). *A study of writing tasks assigned in academic degree programs*. TOEFL Research Report No. 54. Princeton, NJ: Educational Testing Service.
- Halliday, M. A. K. (2005). *Computational and quantitative studies*. London, UK: Continuum.
- Hamp-Lyons, L. & Kroll, B. (1997). *TOEFL 2000—writing: Composition, community, and assessment* (TOEFL Monograph Series, MS 5). Princeton, NJ: Educational Testing Service.
- Hardie, A. & McEnery, T. (2010). On two traditions in corpus linguistics, and what they have in common. *International Journal of Corpus Linguistics* 15(3), p. 384-394.

- Harris, T. & Moreno Jaén, M. (Eds.) (2010). *Corpus linguistics in language teaching*. Bern: Peter Lang.
- Harwood, N. (2005). ‘Nowhere has anyone attempted... In this article I aim to do just that’ A corpus-based study of self-promotional I and we in academic writing across four disciplines. *Journal of Pragmatics*, 37, p. 1207–1231.
- Haswell, R. (1998). Rubrics, prototypes, and exemplars: Categorization theory and systems of writing placement. *Assessing Writing*, 5(2), p. 231-268.
- Hewings, A. (2004). Developing discipline-specific writing: An analysis of undergraduate geography essays. In L. Ravellia & R. Ellis, (Eds.), *Analysing academic writing: Contextualized frameworks*, p. 131-152. London, UK: Continuum.
- Hewings, M. (2006). *Academic writing in context*. Continuum: London.
- Hewings, M. & Hewings, A. (2002). “It is interesting to note that.”: A comparative study of anticipatory ‘it’ in student and published writing. *English for Specific Purposes*, 21(4), p. 367–383.
- Hjortshoj, K. (2001). *The transition to college writing*. Boston, MA: Bedford.
- Hinkel, E. (1997). Indirectness in L1 and L2 academic writing. *Journal of Pragmatics*, 27(3), p. 361-386.
- Hinkel, E. (2001). Matters of cohesion in L1 and L2 academic texts. *Applied Language Learning*, 12, p. 111–132.
- Hinkel, E. (2002a). *Second language writers’ text*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hinkel, E. (2002b). Teaching grammar in writing classes: Tenses and cohesion. In E. Hinkel & S. Fotos (Eds.), *New perspectives on grammar teaching in second language classrooms*, p. 181–198. Mahwah, NJ: Lawrence Erlbaum Associates.

- Hinkel, E. (2003). Simplicity without elegance: Features of sentences in L1 and L2 academic texts. *TESOL Quarterly* 37(2), p. 275-301.
- Hinkel, E. (2004a). *Tense, aspect, and the passive voice in L1 and L2 academic texts*. London: Arnold.
- Hinkel, E. (2004b). *Teaching academic ESL writing*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hinkel, E. (2009). The effects of essay topics on modal verb uses in L1 and L2 academic writing. *Journal of Pragmatics*, 41(4), p. 667–683.
- Hinkel, E. (2011). What research on second language writing tells us and what it doesn't. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning: Volume 2*, p. 523-538. New York, NY: Routledge.
- Hoey, M. (2005). *Lexical priming. A new theory of words and language*. London, UK: Routledge.
- Hoey, M. (1991). *Pattern of lexis in text*. Oxford, UK: Oxford University Press.
- Holmes, J. & Nesi, H. (2009). Verbal and mental processes in academic disciplines. In M. Charles, D. Pecorari, & S. Hunston (Eds.), *Academic writing: At the interface of corpus and discourse*, p. 58-72. London: Continuum International Publishing Group.
- Holmes, J. & Sigley, R. (2002). What's a word like *girl* doing in a place like this? Occasional labels, sexist usage and corpus research. In P. Peters, P. Collins, & A. Smith (Eds.), *New frontiers of corpus research*, p. 247-264. Amsterdam: Rodopi.
- Hubert, M. D. & Bonzo, J. D. (2010). Does second language writing research impact US University foreign language instruction? *System*, 38(4), p. 517-528.
- Hüning, M. (2008). TextSTAT (Simple Text Analysis Tool) 2.8 [Computer Software]. Berlin: Freie Universität. <http://neon.niederlandistik.fu-berlin.de/textstat>

- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hunston, S. (2006). Phraseology and system: A contribution to the debate. In G. Thompson & S. Hunston (Eds.), *System and corpus: Exploring connections*, p. 55-80. London, UK: Equinox.
- Hunston, S. (2011). *Discussion on corpus linguistics as a theory*. [email] (Personal communication, 22 April 2011).
- Hunston, S. & Francis, G. (2000). *Pattern grammar*. Amsterdam: Benjamin Johns Publishing.
- Hunston, S., Francis, G., & Manning, E. (1997). Grammar and vocabulary: Showing the connections. *ELT Journal* 51(3), p. 208-216.
- Hüttner, J. (2008). The genre(s) of student writing: Developing writing models. *International Journal of Applied Linguistics*, 18(2), p. 146-165.
- Hüttner, J. (2010). Purpose-built corpora and student writing. *Journal of Writing Research*, 2(2), p. 197-218.
- Hyland, K. (1997). Is EAP necessary? A survey of Hong Kong undergraduates. *Asian Journal of English Language Teaching*, 7, p. 77-99.
- Hyland, K. (1998). *Hedging in scientific research articles*. Amsterdam: John Benjamins.
- Hyland, K. (1999). Disciplinary discourses: Writer stance in research articles. In C.N. Candlin & K. Hyland (Eds.). *Writing: Texts, processes and practices*, p. 99-121. London, UK: Longman.
- Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, 20, p. 207-226.
- Hyland, K. (2002a). Options of Identity in Academic Writing. *ELT Journal*, 56(4), p. 351-358.

- Hyland, K. (2002b). *Teaching and researching writing*. Essex, England: Pearson Education.
- Hyland, K. (2002c). Activity and evaluation: Reporting practices in academic writing. In J. Flowerdew (Ed.), *Academic discourse*, p. 115-130. Harlow, England: Longman.
- Hyland, K. (2003). *Second language writing*. Cambridge: Cambridge University Press.
- Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing*. Ann Arbor, MI: University of Michigan Press.
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. London, UK: Continuum.
- Hyland, K. (2006). *English for academic purposes—An advanced resource book*. London, UK: Routledge.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), p. 4-21.
- Hyland, K. (2009). Corpus informed discourse analysis: The case of academic engagement. In M. Charles, D. Pecorari, & S. Hunston (Eds.), *Academic writing: At the interface of corpus and discourse*, p. 110-128. London: Continuum International Publishing Group.
- Hyland, K. & Milton, J. (1997). Qualifications and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*, 6, p. 183-205.
- Hyland, K. & Tse, P. (2004). "I would like to thank my supervisor". Acknowledgements in graduate dissertations. *International Journal of Applied Linguistics*, 14(2), p. 259-275.
- Hyland, K. & Tse, P. (2005). Hooking the reader: A corpus study of evaluative *that* in abstracts. *English for Specific Purposes* (24)2, p. 123-139.
- Hyon, S. (1996). Genre in three traditions: Implications for ESL. *TESOL Quarterly*, 30(4), p. 693-722.

- Işık Taşa, E. (2008). A corpus-based analysis of genre-specific discourse of research: The research article and the PhD thesis in ELT. Unpublished PhD Thesis. Middle East Technical University, Ankara.
- Işık Taşa, E. (2010). "In this paper I will discuss...": Current trends in academic writing. *Procedia Social and Behavioral Sciences*, 3, p. 121–126.
- Ismali, S. (2010). Arabic and English persuasive writing of Arabs from a contrastive rhetoric perspective. *ProQuest document ID: 365786902*.
- Jabbour, G. (1997). Where is the writer in a frequency list? Using a corpus of medical research articles in teaching. In B. Lewandowska-Tomaszczyk & P. J. Melia (Eds.), *PALC'97: Practical applications in language corpora*, p. 178-190. Lodz: Lodz University Press.
- Jablonkai, R. (2010). English in the context of European integration: A corpus-driven analysis of lexical bundles in English EU documents. *English for Specific Purposes*, 29(4), p. 253-267.
- James, C. L. (2006). Validating a computerized scoring system for assessing writing and placing students in composition courses. *Assessing Writing*, 11(3), p. 167-178.
- James, M. A. (2010). An investigation of learning transfer in English-for-general-academic-purposes writing instruction. *Journal of Second Language Writing*, 19, p. 183–206.
- Jayne, V., McMartin-Miller, C., Pelaez Morales, C., & Silva, T. (2011). Scholarship on Second Language Writing in 2010: The Year in Review. Presentation at the 45th Annual Teaching English to Speakers of Other Languages (TESOL) Convention, New Orleans, Louisiana, 19 March, 2011.
- Johansson, S. (2007). *Seeing through multilingual corpora. On the use of corpora in contrastive studies*. Amsterdam: John Benjamins.

- Johansson, S., Ebeling, J., & Hofland, K. (1996). Coding and aligning the English-Norwegian parallel corpus. In K. Aijmer, Altenberg, B., & Johansson, M. (Eds.), *Languages in contrast papers from a symposium on text-based cross-linguistic studies*, p. 87-112. Lund: Lund University Press.
- John, S. (2009). Using the Revision Process to Help International Students Understand the Linguistic Construction of the Academic Identity. In M. Charles, D. Pecorari, & S. Hunston (Eds.), *Academic writing: At the interface of corpus and discourse*, p. 272-290. London: Continuum International Publishing Group.
- Johns, A. (1988). The discourse communities' dilemma: Identifying transferable skills for the academic milieu. *English for Specific Purposes*, 7(1), p. 55-59.
- Johns, A. (1995). Teaching classroom and authentic genres: Initiating students into academic cultures and discourses. In D. Belcher & G. Braine (Eds.), *Academic writing in a second language: Essays on research and pedagogy*, p. 277-291. Norwood, NJ: Ablex.
- Johns, A. (1997). *Text, role, and context: Exploring academic literacies*. Cambridge, UK: Cambridge University Press.
- Johns, A. M. & Swales, J. M. (2002). Literacy and disciplinary practices: Opening and closing perspectives. *Journal of English for Academic Purposes*, 1(1), p. 13-28.
- Johnson, D., Malhotra, V., & Vamplew, P. (2006). More Effective Web Search Using Bigrams and Trigrams. *Webology* 3(4). Retrieved from <http://www.webology.ir/2006/v3n4/a35.html>.
- Jordan, R.R. (1997). *English for academic purposes: A guide and resource book for teachers*. Cambridge: Cambridge University Press.
- Jordan, R. R. (2002). The growth of EAP in Britain. *Journal of English for Academic Purposes*, 1, p. 69-78.

- Kaufer, D. S. & Young, R. (1993). Writing in the content areas: Some theoretical complexities. In L. Odell (Ed.), *Theory and practice in the teaching of writing: Rethinking the discipline*, p. 71-104. Carbondale, IL: Southern Illinois University Press.
- Keck, C. (2006). The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *Journal of Second Language Writing* 15(4), p. 261–278.
- Kemppanen, H. (2004). Keywords and ideology in translated history texts: A corpus-based analysis. *Across Languages and Cultures*, 5(1), p. 89–106.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. London, UK: Longman.
- Kennedy, G. (1991). *Between and through: The company they keep and the functions they serve*. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartvik*, p. 95-110. London, UK: Longman.
- Khuwaileh, A. A. (1999). The role of chunks, phrases and body language in understanding co-ordinated academic lectures. *System*, 27(2), p. 249-260.
- Koutsanti, D. (2006). Rhetorical strategies in engineering research articles and research theses: Advanced academic literacy and relations of power. *Journal of English for Academic Purposes*, 5, p. 19–36.
- Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford, UK: Pergamon Press.
- Krashen, S. (1984). Immersion: Why it works and what it has taught us. *Language and Society*, 12, 61-64.
- Krishnamurthy, R. & Kosem, I. (2007). Issues in creating a corpus for EAP pedagogy and research. *Journal of English for Academic Purposes* (6), p. 356–373.

- Kroll, B. (2001). Considerations for teaching an ESL/EFL writing course. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language*, p. 219-232. Boston, MA: Heinle.
- Kroll, B. (2003). *Exploring the dynamics of second language writing*. New York, NY: Cambridge University Press.
- Kroll, K. & Alford, B. (1997). Scholarship, tenure, and composition studies in the two-year college. In R.C. Gebhardt & B.G.S. Gebhardt (Eds.), *Academic advancement in composition studies: Scholarship, publication, promotion, tenure*, p. 57-70. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kusel, P. A. (1992). Rhetorical approaches to the study and composition of academic essays. *System* 20(4), p. 457-469.
- Kwan, B. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes*, 25, p. 30–55.
- Lantolf, J. (2000). *Sociocultural theory and second language learning*. Oxford, UK: Oxford University Press.
- Laviosa, S. (2002). *Corpus-based Translation Studies: Theory, Findings, Applications*. Amsterdam and Atlanta: Rodopi.
- Le, Z. (2003). *Morphix-NLP Live CD Manual*. Retrieved from <http://morphix-nlp.berlios.de/manual/node17.html>.
- Lee, J. (2009). Size matters: An exploratory comparison of small- and large-class university lecture introductions. *English for Specific Purposes*, 28(1), p. 42-57.
- Lee, L. (2010). Exploring wiki-mediated collaborative writing: A case study in an elementary Spanish course. *CALICO Journal*, 27(2), p. 260-276.

- Lee, S. (2010). Command strategies for balancing respect and authority in undergraduate expository essays. *Journal of English for Academic Purposes*, 9, p. 61-75.
- Lee, D. & Swales, J. (2005). A corpus-based EAP course for NNS doctoral students: Moving from available specialised corpora to self-compiled corpora. *English for Specific Purposes* (25), p. 56-75.
- Leech, G. (1992). Corpora and theories of linguistic performance. In J. Svartvik (Ed.), *Directions in corpus linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4–8 August*, p. 105-122. Berlin: Mouton de Gruyter.
- Leech, G., & Svartvik, J. (2002) *A communicative grammar of English*, (3rd ed.). New York: Longman.
- Leki, I. (1990). Potential problems with peer responding in ESL writing classes. *CATESOL Journal*, 3, p. 5-19.
- Leki, I. (2000). Writing, literacy, and applied linguistics. *Annual Review of Applied Linguistics*, 20, 99-115.
- Leki, I. (2007). *Undergraduates in a second language: Challenges and complexities of academic literacy development*. New York: Erlbaum.
- Leki, I. & Carson, J. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly* 28(1), p. 81-101.
- Lester, N., Bertram, C., Erickson, G., Lee, E., Tchako, A., Wiggins, K.D., & Wilson, J. (2003). Writing across the Curriculum: A College Snapshot. *Urban Education* 38(5), p. 5-34.
- Liang, M. Y. (2010). Using synchronous online peer response groups in EFL writing: Revision-related discourse. *Language Learning & Technology*, 14(1), p. 45-65.

- Lillis, T. (1999). Whose “common sense”? Essayist literacy and the institutional practice of mystery. In C. Jones, J. Turner, & B. Street (Eds.), *Students writing in the university: Cultural and epistemological issues*, p. 127–147. Amsterdam: Johns Benjamins.
- Littlewood, W. & Liu, N. F. (1996). *Hong Kong students and their English*. Hong Kong: Macmillan.
- Lin, C. Y. (2010). ‘...that’s actually sort of you know trying to get consultants in...’: Functions and multifunctionality of modifiers in academic lectures. *Journal of Pragmatics* 42, p. 1173–1183.
- Lorenz, G. (1999). Adjective intensification: learners versus native speakers. A corpus study of argumentative writing. *Language and computers: Studies in practical linguistics* 27. Amsterdam: Rodopi.
- Loudermilk, B. C. (2007). Occluded academic genres: An analysis of the MBA Thought Essay. *Journal of English for Academic Purposes*, 6, p. 190–205.
- Low, G.D., Littlemore, J., & Koester, A. (2008). Metaphor use in three UK university lectures. *Applied Linguistics* 29(3), 428-455.
- MacBeth, K. P. (2010). Deliberate false provisions: The use and usefulness of models in learning academic writing. *Journal of Second Language Writing*, 19, p. 33–48.
- MacDonald, S. (1987). Problem definition in academic writing. *College English*, 49, p. 315-331.
- Marchenkova, L. (2005). Language, culture, and self: The Bakhtin-Vygotsky encounter. In J. Hall, G. Vitanova, & L. Marchenkove (Eds.), *Dialogue with Bahktin on second and foreign language learning*, p. 171-188. Mahwah, NJ: Lawrence Erlbaum Associates.
- Marco, M. J. L. (2000). Collocational frameworks in medical research papers: a genre-based study. *English for Specific Purposes*, 19(1), p. 63-86.

- Martinez, I. A. (2001). Impersonality in the research article as revealed by analysis of the transitivity structure. *English for Specific Purposes* 20(3), p. 227-247.
- Mason, O. (2010). *Running the NAFWiC files through the noun extractor*. [email] (Personal communication, 22 September 2010).
- Mason, O. & Tufis, D. (1997). 'Probabilistic Tagging in a Multi-lingual Environment: Making an English Tagger Understand Romanian'. *Proceedings of the Third International TELRI Seminar*, Montecatini, October 1997, p.165-168.
- Matsuda, P. K. (1998). Situating ESL writing in a cross-disciplinary context. *Written Communication*, 15(1), p. 99-121.
- Matsuda, P. K. (1999). Composition studies and ESL writing: A disciplinary division of labor. *College Composition and Communication*, 50(4), p. 699-721.
- Matsuda, P. K. (2003). Second language writing in the twentieth century: A situated historical perspective. In Kroll, B. (Ed.), *Exploring the dynamics of second language writing*, p. 15-34. Cambridge, UK: Cambridge University Press.
- Matsuda, P. K., Canagarajah, A., Harklau, L., Hyland, K., Warschauer, M. (2003). Changing currents in second language writing research: A colloquium. *Journal of Second Language Writing* 12, p. 151-179.
- McCarthy, L. (1987). A Stranger in Strange Lands: A College Student Writing Across the Curriculum. *Research in the Teaching of English*, 21, p. 233-65.
- McCarthy, M., McCarten, J., & Sandiford, H. (2005). *Touchstone 1*. New York, NY: Cambridge University Press.
- McKay, S. (1994). Developing ESL writing materials. *System*, 22(2), p. 195-203.

- McEnery, T. & Xiao, R. (2011). What Corpora Can Offer in Language Teaching and Learning. In: Hinkel, E. (ed). *Handbook of Research in Second Language Teaching and Learning: Volume 2*. New York, NY: Routledge.
- McEnery, T. & Wilson, A. (2001). *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies*. New York, NY: Routledge.
- Meyer, C. F. (2002). *English corpus linguistics: An introduction*. Cambridge, UK: Cambridge University Press.
- Mittan, R. (1989). The peer review process: Harnessing students' communicative power. In D.M. Johnson & D.H. Roen (Eds.), *Richness in writing: Empowering ESL students*, p. 207-219. New York, NY: Longman.
- Moghtader, M., Cotch, A., & Hague, K. (2001). The First-Year Composition Requirement Revisited: A Survey. *College Composition and Communication*, 52(3), p. 455-467.
- Moon, R. (1998). Frequencies and Forms of Phrasal Lexemes in English. In A. Cowie (Ed.), *Phraseology. Theory, analysis and applications*, p. 79-100. Oxford, UK: Oxford University Press.
- Moore, T. & Morton, J. (2010). Dimensions of difference: A comparison of university writing and IELTS writing. *Journal of English for Academic Purposes*, 4, p. 43-66.
- Morris, K. (2000). The service myth. Why freshman composition doesn't serve "us" or "them." In R. Wallace, A. Jackson, and S.L. Wallace (Eds.), *Reforming college composition: Writing the wrongs*, p. 113-122. Westport, CT: Greenwood Press.
- Mukherjee, J. (2005). *English ditransitive verbs: Aspects of theory, description and a usage-based model*. Amsterdam: Rodopi.

- Mukherjee, J. (2010). Corpus linguistics versus corpus dogmatism—*pace* Wolfgang Teubert. *International Journal of Corpus Linguistics*, 15(3), p. 370–378.
- Mukherjee J. & Rohrback, J.M. (2006). Rethinking applied corpus linguistics from a language-pedagogical perspective: New departures in learner corpus research. In B. Ketterman & G. Marko, (Eds.), *Planning, gluing, and painting corpora: Inside the applied corpus linguist's corpus workshop*, p. 205-232. Frankfurt: Peter Lang.
- Mungra, P. & Webber, P. (2010). Peer review process in medical research publications: Language and content comments. *English for Specific Purposes*, 29(1), p. 43-53.
- National Center for Education Statistics, U.S. Department of Education. (2010). *The Condition of Education 2010* (NCES 2010-028), Indicator 43.
- Nesi, H. (2009). Extended abstract: A multidimensional analysis of student writing across levels and disciplines. In: Edwardes, M. (Ed), *Taking the Measure of Applied Linguistics: Proceedings of the BAAL Annual Conference*, University of Swansea, 11-13 September 2008. London: BAAL/Scitsiugnil Press.
- Nesi, H. & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics* 11(3), p. 147-168.
- Nassaji, H. & Cumming, A. (2000). What's in a ZPD? A case study of a young ESL student and teacher interacting through dialogue journals. *Language Teaching Research* 4(2), p. 95-121.
- Nesselhauf, N. & Römer, U. (2007). Lexical-grammatical patterns in spoken English: The case of the progressive with future time reference. *International Journal of Corpus Linguistics*, 12(3), p. 297–333.
- Norris, J. M. & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning* 50(3), p. 417-528.

- North, S. (2005). Disciplinary variation in the use of theme in undergraduate essays. *Applied Linguistics* 26(3), p. 431–452.
- O'Brien, T. (1995). Rhetorical structure analysis and the case of the inaccurate, incoherent source-hopper. *Applied Linguistics*, 6, p. 442–482.
- O'Keefe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom*. Cambridge: Cambridge University Press.
- Ong, J. & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*, 19(4), p. 218-233.
- Oshima, A. & Hogue, A. (2006). *Writing academic English*. White Plains, NY: Pearson Longman.
- Oshima, A. & Hogue, A. (2007). *Introduction to academic writing*. White Plains, NY: Pearson Longman.
- Panofski, M., Smith, S., Santos, J., Fogelman, C., Harrington, M., Kenney, E. (2005). *Approaches to writing instruction for adolescent English language learners*. Providence, RI: The Education Alliance at Brown University.
- Parrish, B. (2009). *Four point listening and speaking 2*. Ann Arbor, MI: University of Michigan Press.
- Partington, A. (1998). *Patterns and meanings*. Amsterdam: John Benjamins.
- Peacock, M. (2011). A comparative study of *introductory it* in research articles across eight disciplines. *International Journal of Corpus Linguistics* 16(1), p. 72–100.
- Pearson Longman. (2010). Longman English language teaching catalogue. [online] Available at <<http://elcatalogue.pearsoned->

ema.com/Course.asp?Callingpage=Catalogue&CourseID=GC> [Accessed 5 December 2010].

Pearson Longman. (2011). Longman Corpus Network. [online] Available at <<http://www.pearsonlongman.com/dictionaries/corpus/index.html>> [Accessed 15 February 2011].

Pecorari, D. (2009). Formulaic Language in Biology: A Topic-specific Investigation. In M. Charles, D. Pecorari, & S. Hunston (Eds.), *Academic writing: At the interface of corpus and discourse*, p. 91-104. London: Continuum International Publishing Group.

Petraglia, J. (1995). *Reconceiving writing rethinking writing instruction*. Mahwah, NJ: Laurence Erlbaum.

Polio, C. (2003). Research on second language writing: An overview of what we investigate and how. In B. Kroll (Ed.), *Exploring the dynamics of second language writing*, p. 35-66. New York, NY: Cambridge University Press.

Qin, J. & Karabacak, E. (2010). The analysis of Toulmin elements in Chinese EFL university argumentative writing. *System*, 38(3), p. 444-456.

Quirk, R., Greenbaum, S., Leech, G., Svartvik, J., (1985). *A comprehensive grammar of the English language*. Longman, New York.

Radia, P. & Stapleton, P. (2010). Unconventional sources as a new convention: The shifting paradigm of undergraduate writing. *Internet and Higher Education*, 12(3-4), p. 156-164.

Raimes, A. (1991). Out of the woods: Emerging traditions in the teaching of writing. *TESOL Quarterly*, 25(3), p. 407-430.

- Ravelli, L. J. (2004). Signalling the organization of written texts: Hyper-themes in management and history essays. In L. J. Ravelli & Ellis, R. (Eds.), *Analysing Academic Writing: Contextualized Frameworks*, p. 104-125. London: Continuum.
- Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4), p. 519-549.
- Reid, J. (2001). Advanced EAP writing and curriculum design: What do we need to know? In T. Silva & P.K. Matsuda (Eds.), *On second language writing*, p. 143-156. Mahwah, NJ: Lawrence Erlbaum Associates.
- Reid, J. (2008). *Writing myths*. Ann Arbor: University of Michigan Press.
- Reza Rezaei, A. & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, 15(1), p. 18-39.
- Rhee, E. H. (2010). Complexities and dynamics of Korean graduate students' textual borrowing in academic writing. *ProQuest document ID: 607931638*.
- Rob, T. (2003). Google as a Quick 'n Dirty Corpus Tool. *TESOL-EJ*, 7(2).
- Robinson, M., Stoller, F., Costanza-Robinson, M., & Jones, J.K. (2008). *Write like a chemist: A guide and resource*. Oxford: Oxford University Press.
- Robson, J., Francis, B., & Read, B. (2002). Writes of passage: Stylistic features of male and female undergraduate history essays. *Journal of Further and Higher Education*, 26(4), p. 351-362.
- Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics*, 7, p. 141-163.
- Römer, U & Schulze, R. (Eds.). (2009). *Exploring the lexis–grammar interface*. Amsterdam: John Benjamins.

- Römer, U. & Wulff, S. (2010). Applying corpus methods to writing research: Explorations of MICUSP. *Journal of Writing Research*, 2(2), p. 99-127.
- Russell, D. (1995). Activity Theory and Its Implications for Writing Instruction. In J. Petraglia (Ed.), *Reconceiving writing rethinking writing instruction*, p. 51-77. Mahwah, NJ: Laurence Erlbaum.
- Salehzadeh, J. (2005). *Academic listening strategies*. Ann Arbor, MI: University of Michigan Press.
- Samraj, B. (2004). Discourse features of the student-produced academic research paper: variations across disciplinary courses. *Journal of English for Academic Purposes* (3), p. 5–22.
- Samraj, B. (2008). A discourse analysis of master's theses across disciplines with a focus on introductions. *Journal of English for Academic Purposes*, 7, p. 55-67.
- Schmitt, D. & Schmitt, N. (2005). *Focus on vocabulary*. White Plains, NY: Pearson Education.
- Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Mahwah, NJ: Lawrence Erlbaum.
- Schleppegrell, M. (2002). Challenges of the science register for ESL students: Errors and meaning-making. In M. Schleppegrell & M. Colombi (Eds.), *Developing advanced literacy in first and second languages* (pp. 119–142). Mahwah, NJ: Lawrence Erlbaum Associates.
- Schleppegrell, M. J., Achugar, M., & Orteiza, T. (2004). The grammar of history: Enhancing content-based instruction through a functional focus on language. *TESOL Quarterly*, 38(1), p. 67–93.

- Schleppegrell, M. J., & Colombi, M. C. (Eds.). (2002). *Developing advanced literacy in first and second languages*. Mahwah, NJ: Lawrence Erlbaum.
- Schlitz, S. A. (2010). Introduction to special issue: Exploring corpus-informed approaches to writing research. *Journal of Writing Research*, 2(2), p. 91- 98.
- Schmid, L. M. (1999). *The effects of peer response on essay drafts*. MA thesis. Sacramento: California State University.
- Schönefeld, D. (1999). Corpus linguistics and cognitivism. *International Journal of Corpus Linguistics*, 4(1), p.131-171.
- Scott, M. (2000). Focusing on the text and its key words. In L. Burnard & T. McEnery (Eds.), *Rethinking language Pedagogy from a corpus perspective*, Vol. 2, p. 103-122. Frankfurt: Peter Lang.
- Scott, M. & Tribble, C. (2006). Textual patterns: Key words and corpus analysis in language education. Amsterdam: John Benjamins.
- Seal, B. (2010). *Writing Textbooks*. [email] (Personal communication, 7 October 2010).
- Shaw, P. (2009). Linking adverbials in student and professional writing in literary studies: What makes writing mature. In Charles, M., Pecorari, D., & Hunston, S. (Eds.), *Academic writing: At the interface of corpus and discourse*, p. 215-235. London: Continuum International Publishing Group.
- Shaw, P. & Liu, E. (1998). What develops in the development of second language writing? *Applied Linguistics*, 19(2), p. 225-254.
- Sheldon, E. (2009). From one I to another: Discursive construction of self-representation in English and Castilian Spanish research articles. *English for Specific Purposes*, 28, p. 251–265.

- Silva, T. (1990). Second language composition instruction: Developments, issues, and directions in ESL. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom*, p. 11-23. Cambridge: UK: Cambridge University Press.
- Silva, T. (1997). Differences in ESL and native-English-speaker writing: The research and its implications. In C. Severino, J. C. Guerra, & J. E. Butler (Eds.), *Writing in multicultural settings*, p. 209-219. New York, NY: The Modern Language Association of America.
- Silva, T., & Leki, I. (2004). Family matters: The influence of applied linguistics and composition studies on second language writing studies—past, present, and future. *Modern Language Journal*, 88, 1-13.
- Silver, M. (2003). The stance of stance: a critical look at ways stance is expressed and modeled in academic discourse. *Journal of English for Academic Purposes*, 2, p. 359–374.
- Simpson, R. & Swales, J. (2001). *Corpus linguistics in North America*. Ann Arbor: University of Michigan Press.
- Sinclair, J. (1990). *Collins COBUILD English grammar*. London: Harper Collins.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford, UK: Oxford University Press.
- Sinclair, J. (Ed.). (2004). *How to use corpora in language teaching*. Amsterdam: John Benjamins Publishing.
- Sippel, K. (2010). *Writing books*. [email] (Personal communication, 3 October 2010).
- Soles, D. (2005). An analysis of the style of exemplary first-year writing. *Teaching English in the Two Year College*, 33(1), p. 38-49.

- Sotillo, S. & Starace-Natasi, D. (1999). Political discourse of a working-class town. *Discourse and Society*, 10(2), p. 249-276.
- Spack, R. (1998). Initiating ESL students into the academic discourse community: How far should we go? In V. Zamel & R. Spack, *Negotiating academic literacies*, p. 85-104. Mahwah, NJ: Lawrence Erlbaum.
- Starfield, S. (2004). Word power: negotiating success in a first-year sociological essay. In L. J. Ravelli & R. Ellis (eds.), *Analysing academic writing: Contextualized frameworks*, p. 66-83. New York: Continuum.
- Starfield, S. & Ravelli, L. J. (2006). “The writing of this thesis was a process that I could not explore with the positivistic detachment of the classical sociologist”: Self and structure in New Humanities research theses. *Journal of English for Academic Purposes*, 5, p. 222–243.
- Stoller, F. L. (1994). The Diffusion of Innovations in Intensive ESL Programs. *Applied Linguistics* 15(3), p. 300-327.
- Strauch, A. (2005). *Writers at work: The short composition*. New York, NY: Cambridge University Press.
- Stubbs, M. (1993). British traditions in text analysis: From Firth to Sinclair. In M. Baker, F. Francis & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair*, p. 1-46. Amsterdam: John Benjamins.
- Stubbs, M. (1995). Collocation and semantic profiles: On the cause of the trouble with quantitative methods. *Function of Language*, 2(1), p. 1-33.
- Stubbs, M. (2001). *Words and phrases: corpus studies of lexical semantics*. Oxford: Blackwell Publishers.

- Stubbs, M. (2006). Corpus analysis: The state of the art and three types of unanswered questions. In G. Thompson & S. Hunston (Eds.), *System and corpus: Exploring connections*, p. 15-36. London, UK: Equinox.
- Stubbs, M. (2010). Three concepts of keywords. In M. Bondi & M. Scott, (Eds.), *Keyness in texts*, p. 21-42. Amsterdam: John Benjamins.
- Stubbs, M. & Barth, I. (2003). Using recurrent phrases as text-type discriminators: A quantitative method and some findings. *Functions of Language*, 10(1), p. 61-104.
- Suh, B. R. (2010). Written feedback in second language acquisition: Exploring the roles of type of feedback, linguistic targets, awareness, and concurrent verbalization. *ProQuest document ID: 275981651*.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. New York, NY: Cambridge University Press.
- Swales, J. (1995). The Role of the textbook writing research. *English for Specific Purposes*, 14(1), p. 3-18.
- Swales, J. & Feak, C. (2006). Reflections on collaborate practice in EAP materials production. In M. Hewings, *Academic writing in context*, p. 215-226. Continuum: London.
- Swales, J. & Feak, C. (2009). *Academic writing for graduate students: Essential tasks and skills*. Ann Arbor, MI: University of Michigan Press.
- Taiwo, R. (2010). *Collocation in non-native English: A study of Nigerian ESL writing*. Saarbrucken, Germany: VDM Verlag Dr. Muller.
- Teubert, W. (2005). "My version of corpus linguistics". *International Journal of Corpus Linguistics*, 10(1), p. 1-13.

- Teubert, W. & Cermakova, A. (2007). *Corpus linguistics: A short introduction*. London, UK: Continuum.
- Thompson, G. & Hunston, S. (Eds.). *System and corpus: Exploring connections*. London, UK: Equinox.
- Tobash, L. (2005). Choosing a college major. [online] American.gov—Engaging the world. Available at <<http://www.america.gov/st/educ-english/2008/April/20080518234235SrenoD0.9282038.html>> [Accessed 9 July 2010].
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins Publishing.
- Tomasello, M. (2003). *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Tomlinson, B. (1998). *Materials development in language teaching*. Cambridge, UK: Cambridge University Press.
- Tribble, C. (2010). The game I'm interested in. *International Journal of Corpus Linguistics*, 15(3), p. 359–362.
- Tribble, C. (2009). Writing academic English—a survey review of current published resources. *ELT Journal*, 63(4), p. 400-417.
- Tribble, C. (2002). Corpora and corpus analysis: New windows on academic writing. In J. Flowerdew (Ed.), *Academic discourse*, p. 131-149. Harlow, England: Longman.
- Tribble, C. (2001). Small corpora and teaching writing. In M. Ghadessy, A. Henry, & R. L. Roseberry (Eds.), *Small corpus studies and ELT*, p. 381–408. Amsterdam: John Benjamins Publishing.
- Tribble, C. (1999). Genres, keywords, teaching: Towards a pedagogic account of the language of project proposals. In L. Burnard & T. McEnery (Eds.), *Rethinking*

- language pedagogy from a corpus perspective*, p. 75-90. Frankfurt: Peter Lang Publishers.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46, p. 327-369.
- Truscott, J. (1999). The case for “The Case Against Grammar Correction in L2 Writing Classes”: A response to Ferris. *Journal of Second Language Writing*, 8(2), p. 111-122.
- Truscott, J. (2010). Some thoughts on Anthony Bruton’s critique of the correction debate. *System*, 38(2), p. 329-335.
- Tsui, A. (2005). ESL teachers’ questions and corpus evidence. *International Journal of Corpus Linguistics* (10), p. 335-356.
- UCIEP. University and College Intensive English Programs. [online] Available at <www.uciep.org> [Accessed 9 July 2010].
- University of Illinois at Chicago. (2008). *Purpose of the general education program*. [online] Available at <http://www.uic.edu/depts/oa/gened/purpose.html> [Accessed 9 July 2010].
- Vassileva, I. (2001). Commitment and detachment in English and Bulgarian academic writing. *English for Specific Purposes*, 20, p. 83-102.
- Vongpumivitch, K., Huang, J., & Chang, Y.C. (2009). Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes*, 28, p. 33–41.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

- Walker, C. (2011). How a corpus-based study of the factors which influence collocation can help in the teaching of business English. *English for Specific Purposes*, 30(2), p. 101-112.
- Walvoord, B. (1985). Freshman, 'Focus,' and Writing Across the Curriculum. *Freshman English News* 14(2), p. 13 17.
- Walvoord, B & McCarthy, L. (1990). *Thinking and writing in College: A naturalistic study of students in four disciplines*. Urbana, IL: National Council of Teaching of English.
- Ward, J. (2007). Collocation and technicality in EAP engineering. *Journal of English for Academic Purposes*, 6(1), p. 18-35.
- Wardle, E. (2009). “Mutt Genres” and the Goal of FYC: Can We Help Students Write the Genres of the University? *College Composition and Communication*, 60(4), p. 765-789.
- Warschauer, M. (2010). Invited commentary: New tools for teaching writing: *Language Learning and Technology*, 14(1), p. 3-8.
- Wells L. & Valcourt, G. (2010). *Vocabulary mastery series*. Ann Arbor, MI: University of Michigan Press.
- White, L. (1987). Against comprehensible input: The input hypothesis and the development of second language competence. *Applied Linguistics*, 8, p. 95-110.
- Williams, G. (2010). Many rooms with corpora. *International Journal of Corpus Linguistics*, 15(3), p. 400–407.
- Williams, G. (1998). Collocational networks: Interlocking patterns of lexis in a corpus of plant biology research articles. *International Journal of Corpus Linguistics*, 3(1), p. 151-171.
- Williams, I.A. (1996). A contextual study of lexical verbs in two types of medical research report: Clinical and experimental. *English for Specific Purposes*, 15, p. 175-197.

- Wulff, S. & Römer, U. (2009). Becoming a proficient academic writer: Shifting lexical preferences in the use of the progressive. *Corpora*, 4(2), p. 115-133.
- Wynne, M. (2010). Interdisciplinary relationships. *International Journal of Corpus Linguistics*, 15(3), p. 425–427.
- Xie, Y., Ke, F., & Sharma, P. (2008). The effect of peer feedback for blogging on college students' reflective learning processes. *Internet and Higher Education*, 11, p. 18–25.
- Yancey, K. & Huot, B. (Eds.). (1997). *Assessing writing across the curriculum: Diverse approaches and practices*. Greenwich, CT: Ablex.
- Yancey, K. & Weiser, L (Eds.). (1997). *Situating portfolios: Four perspectives*. Logan, UT: Utah State University Press.
- Yoo, I. (2008). A corpus analysis of (the) last/next + temporal nouns. *Journal of English Linguistics*, 36(1), p. 39-61.
- Zemach, D. & Stafford-Yilmaz, L. (2008). *Writers at work: The essay*. New York, NY: Cambridge University Press.
- Zhang, X. M. & Xhang, X. H. (2010). Thematic progression and discourse coherence of college English writing. *US-China Foreign Language*, 8(8), p. 31-39.
- Zhao, H. (2010). Investigating learners' use and understanding of peer and teacher feedback on writing: A comparative study in a Chinese English writing classroom. *Assessing Writing*, 15(1), p. 3-17.

APPENDIX 1 SAMPLE GENERAL EDUCATION COURSE REQUIREMENTS

Appendix 1A GenEd requirements from University of California, Davis

(<http://registrar.ucdavis.edu/UCDwebcatalog/ugraded/gereqt.html>)

The General Education (GE) requirement promotes the intellectual growth of all undergraduates by ensuring that they acquire a breadth of knowledge that will enlarge their perspectives beyond the focus of a major and serve them well as participants in a knowledge-based society. It seeks to stimulate continued intellectual growth by providing students with knowledge not only of the content but also of the methodologies of different academic disciplines. It involves students in the learning process by its expectation of considerable writing and class participation. It encourages students to consider the relationships between disciplines.

The GE requirement has three components: Topical Breadth, Social-Cultural Diversity and Writing Experience.

Topical Breadth – 6 courses

A GE course in topical breadth addresses broad subject areas that are important to the student's general knowledge. It takes a critical, analytical perspective on knowledge, considering how knowledge has been acquired and the assumptions, theories, or paradigms that guide its use. Topical breadth courses are grouped into three broad subject areas of knowledge, Arts and Humanities, Science and Engineering, and Social Sciences.

Three selected courses in each of the two subject areas other than your major's assigned area:

Arts and Humanities. Courses in this area provide students with knowledge of significant intellectual traditions, cultural achievements and historical processes.

African American Studies	Film Studies
American studies	French
Anthropology	German
Art History	History
Art Studio	Human Development
Asian American Studies	International Agriculture Development
Chicano Studies	Italian
Chinese	Japanese
Classics	Jewish Studies
Comparative Literature	Linguistics
Critical Theory	Medieval Studies
Design	Middle East/S. Asian Studies
Dramatic Art	Music
East Asian Studies	Native American Studies
Economics	Nutrition
English	Philosophy
Environmental Science and Policy	Political Science
Exercise Biology	Psychology

Religious Studies
Russian
Science and Society
Sociology

Spanish
Textiles and Clothing
Women's Studies

Science and Engineering. Courses in this area provide students with knowledge of major scientific ideas and applications. They seek to communicate the scope, power, limitations and appeal of science.

Animal Science
Anthropology
Atmospheric Science
Avian Sciences
Biological Sciences
Biotechnology
Chemistry
Chicano Studies
Design
Engineering
Entomology
Environmental Horticulture
Environmental and Resource Science
Environmental Science and Policy
Environmental Toxicology
Evolution and Ecology
Exercise Biology
Food Science and Technology
Geology

Hydrologic Science
International Agricultural Development
Landscape Architecture
Math and Physical Science
Mathematics
Microbiology
Molecular and Cellular Biology
Nature and Culture
Neuro, Physio, and Behavior
Nutrition
Physics
Plant Biology
Plant Pathology
Plant Science
Science and Tech Studies
Science and Society
Soil Science
Viticulture and Enology
Wildlife, Fish, and Conservation Biology

Social Sciences. Courses in this area provide students with knowledge of the individual, social, political and economic activities of people.

African American Studies
Agriculture and Resource Economy
Agricultural Education
American studies
Animal Science
Anthropology
Asian American Studies
Chicano Studies
Communication
Communication and Reg Development
Consumer Sciences
Dramatic Art
Economics
Education
Entomology
Environmental Science and Policy
Exercise Biology
Fiber and Polymer Science

Food Science and Technology
Geology
History
Human Development
Hydrologic Science
International Agriculture Development
Jewish Studies
Landscape Architecture
Linguistics
Native American Studies
Nutrition
Philosophy
Physical Education
Political Science
Psychology
Religious Studies
Science and Technology Studies
Science and Society

Sociology
Spanish

Textiles and Clothing
Women's Studies

Social-Cultural Diversity – 1 course

GE courses in social-cultural diversity teach students the significance of the many patterned differences that characterize human populations—particularly differences of gender, race, ethnicity, sexuality, religion or social class.

One selected course from the following departments:

African American Studies
Agriculture and Economics
American studies
Anthropology
Art History
Asian American Studies
Chicano Studies
Chinese
Comparative Literature
Design
Dramatic Art
East Asian Studies
English
Film Studies
French
German
Greek
History
Italian
Japanese
Jewish Studies
Landscape Architecture
Latin
Linguistics
Medieval Studies
Middle East/S. Asian Studies
Music
Native American Studies
Nature and Culture
Philosophy
Religious Studies
Russian
Science and Society
Spanish
Technocultural Studies
Women's Studies

Writing Experience – 3 courses

GE courses in writing experience improve student writing through instruction and practice. Writing assignments are designed to encourage students to think critically and communicate effectively.

Three selected courses which require one extended writing assignment (five pages or more) or multiple short assignments. Writing is evaluated not only for content, but also for organization, style, use of language, and logical coherence.

Appendix 1B GenEd requirements from Arkansas Tech University

(<http://www.atu.edu/academics/catalog/graduation-requirements.html#GenEdRequirements>)

The general education curriculum is designed to provide a foundation for knowledge common to educated people and to develop the capacity for an individual to expand that knowledge over his or her lifetime. Students who have completed the general education curriculum at Arkansas Tech University will be able to:

Communicate effectively
Think critically
Develop ethical perspectives
Apply scientific and quantitative reasoning
Demonstrate knowledge of the arts and humanities
Understand wellness concepts

To accomplish the above goals, Arkansas Tech requires the completion of the following general education curriculum. Students should refer to the curriculum in their major area of study for specific courses either recommended or required by the academic department to fulfill the general education requirements.

English - 6 hours

Three hours from one of the following:

ENGL 1013 Composition I
ENGL 1043 Honors Composition I

Three additional hours from one of the following:

ENGL 1023 Composition II
ENGL 1053 Honors Composition II

Mathematics - 3 hours

Three hours from one of the following:

MATH 1003 College Mathematics
MATH 1113 College Algebra
Any higher level mathematics course

Science - 8 hours

Complete two of the following, for a total of eight hours of science (four hours of biological sciences and four hours of physical sciences required for graduation):

A. BIOL 1014* Introduction to Biological Science OR any other biology course (BIOL) that includes a lab

B. PHSC 1013* Introduction to Physical Science and PHSC 1021* Physical Science Laboratory OR any other physical science course (CHEM, GEOL, PHYS, PHSC) that includes a lab

C. BIOL/PHSC 1004* Principles of Environmental Science (course may be taken one time and will count for either biological sciences or physical sciences credit)

*Note that the science courses marked above are designed to meet general education objectives.

Physical Activity - 2 hours

Two hours from the following:

Physical education activity courses

Recreation (RP) coeducational activity courses

Wellness science activity courses

Theatrical dance activity

Appropriate military science courses completed through cross-enrollment agreement with UCA.

Fine Arts - 3 hours

Three hours from one of the following:

*ART 2123 Experiencing Art

MUS 2003 Introduction to Music

TH 2273 Introduction to Theatre

*ENGL 2173 Introduction to Film

*JOUR 2173 Introduction to Film

Art Majors:

Art Education Majors Take ART 2123

Fine Arts and Graphic Design majors take any of the above options except ART 2123

Music Majors:

Any of the above course options except MUS 2003

Humanities - 3 hours

Three hours from one of the following:

*ENGL 2003 Introduction to World Literature

ENGL 2013 Introduction to American Literature

ENGL 2023 Honors World Literature

PHIL 2003 Introduction to Philosophy

PHIL 2043 Honors Introduction to Philosophy

Social Sciences - 12 hours

Three hours from one of the following:

HIST 1903 Survey of American History
POLS 2003 American Government

Nine additional hours from the following:

*HIST 1503 World Civilization I
*HIST 1513 World Civilization II
HIST 1543 Honors World Civilization I
HIST 2003 U.S. History I
HIST 2013 U.S. History II
HIST 2043 Honors U.S. History I
HIST 1903 Survey of American History
POLS 2003 American Government
ECON 2003 Principles of Economics I
ECON 2103 Honors Principles of Economics I
SOC 1003 Introductory Sociology
PSY 2003 General Psychology
*ANTH 1213 Introduction to Anthropology OR
*ANTH 2003 Cultural Anthropology
*GEOG 2013 Regional Geography of the World
AMST 2003 American Studies

*Of the above 18 hours in Fine Arts, Humanities, and Social Science, three hours must be from one of the following:

ART 2123 Experiencing Art
ENGL 2173 Introduction to Film
JOUR 2173 Introduction to Film
ENGL 2003 Introduction to World Literature
HIST 1503 World Civilization I
HIST 1513 World Civilization II
ANTH 1213 Introduction to Anthropology OR
ANTH 2003 Cultural Anthropology
GEOG 2013 Regional Geography of the World

APPENDIX 2
NAFWiC INFORMED CONSENT FORM

University of Birmingham, UK
Department of English
Informed Consent

Title: A Lexico-Grammatical Profile of Freshman Composition Writing in North America

Principal Investigator: Gena Bennett

I. Purpose

You are invited to participate in a research study. The purpose of the study is to investigate the most frequent and useful linguistic features of successful writing assignments from freshman composition courses at institutions of higher education in the United States. Approximately 500 participants from each institution will be recruited for this study. Participation will require approximately 30 minutes of your time over the course of one or two days.

II. Procedures

If you decide to participate in this study, you will first complete a consent form which will remain in a sealed envelope. Writing assignments from your freshman composition course will be submitted for this study. The researcher will compile all submissions into a corpus which will be queried for the most prominent and salient linguistic features. The results will inform second language writing materials for college-bound English language learners.

III. Risks

In this study, you will not have any more risks than you would in a normal day of life.

IV. Benefits

Participation in this study may not benefit you personally. Overall, we hope to gain information about successful freshman composition writing assignments to create better materials to help future college-bound English language learners.

V. Voluntary Participation and Withdrawal

Participation in research is voluntary. You may submit any number of assignments you wish.

VI. Confidentiality:

The researcher will keep your records private to the extent allowed by law. Your name will not be on any submitted writing assignments. Your signed consent form will be kept in a

sealed envelope in a locked cabinet. Your work will be used for research and teaching purposes, including published research and writing activities, but no names will be recorded and every effort will be made never to make public anything which could identify the writer of any essay. All findings in this study will be summarized and reported in group form.

VII. Contact Persons:

Contact Gena Bennett at [REDACTED], or speak with your department head if you have questions about this study.

If you are willing to volunteer for this research, please sign below.

Participant

Date

APPENDIX 3 QTAG ACCURACY ANALYSIS

TextSTAT 2.8e
(c) Matthias Häring 2000/2009

values_NNS (verb)
heritage_NN
literacy_NN
His_PP\$ purpose_NN
literacy_NN ._.
audience_NN
mobility_NN
audience_NN
needs_NNS (verb)
topic_NN
text_NN
books_NNS
vocabulary_NN
paragraph_NN
paragraph_NN
fence_NN
Example_NN
text_NN
childhood_NN
years_NNS
boy_NN
family_NN
misconceptions_NNS
person_NN
group_NN
culture_NN
stereotypes_NNS
Example_NN
trials_NNS
periods_NNS
narrative_NN
picture_NN
words_NNS
door_NN
narrative_NN
world_NN
children_NNS
school_NN
expectations_NNS
teacher_NN
answers_NNS
volunteers_NNS

children_NN
society_NN
racism_NN
society_NN
world_NN
pathos_NN
ethos_NN
appeals_NN
lack_NN
ethos_NN
appeals_NNS
opportunity_NN
resources_NNS
people_NN
literacy_NN
solutions_NNS
change_NN (verb)
attitudes_NN
example_NN
Example_NN
citation_NN
paychecks_NNS
hope_NN
fear_NN
government_NN
surplus_NN
opportunity_NN
evaluation_NN
people_NN
way_NN
people_NN
Ethnicity_NN
intelligence_NN
opportunity_NN
evaluation_NN
people_NN
way_NN
people_NN
conclusion_NN
structures_NNS
structures_NNS
person_NN
expectations_NNS
people_NN
people_NN
description_NN
type_NN
evidence_NN

text_NN
stages_NNS
life_NN
stereotypes_NNS
childhood_NN
student_NN
adult_NN
structure_NN
passage_NN
lessons_NNS

APPENDIX 4

LINGUISTIC FEATURES OF SCHOLARLY WRITING BY PART OF SPEECH

Nouns

Nouns are the most frequent word class in academic prose. 60% of all content words in academic prose are nouns (Biber, et al., 1999, p. 65).

Singular nouns are most common in all registers; however, plural nouns are used more in academic prose than in other registers (Biber, et al., 1999, p. 291).

Academic prose uses plural nouns 3-4 times more than conversation.

Derived nouns are most common in academic prose (Biber, et al., 1999, p. 322).
particularly *-tion*.

60% of all noun phrases have modifiers (Biber, et al., 1999, p. 578).

25% are pre modifiers (30% of pre modifiers are nouns— Four premodifying nouns are extremely productive and extremely frequent: *government, police, home, world*. Twelve premodifying nouns are extremely productive and relatively frequent: *business, car, city, council, family, health, labor, market, party, record, security, TV*) (Biber, et al., 1999, p. 589).

79-80% of pre modifiers are one word; ~20% two words (most of these are *adj+and+adj*); 1% 3+ words (Biber, et al., 1999, p. 597).

20% are post modifiers

12% are both

~95+% head nouns are common nouns; they take both pre and post modifiers (Biber, et al., 1999, p. 580). *One* and *those* can also be head nouns. *One* sometimes takes pre modifiers, but *other* takes post modifiers; *those* does not take pre modifiers, but often takes post modifiers

Post modified and complemented noun phrases are extremely frequent in academic English (Carter & McCarthy, 2006, p. 269).

Post modification by non-finite clauses is frequent in academic style (Carter & McCarthy, 2006, p. 270).

Full relative clauses as well as *to*-clauses are rare in academic prose (Biber, et al., 1999, p. 637).

Noun complement clauses are only moderately common in academic prose (compared to post modifiers) (Biber, et al., 1999, p. 647).

that, to, -ing clauses are evenly distributed across academic prose, most use with definite singular nouns.

Wh- clauses never used in academic prose

25% are extraposed.

Retention of *that* is normal in academic prose (omission is rare).

Greater than 30% of stance markings in academic prose are in complement clauses.

Definite noun phrases are used in anaphoric expressions in academic prose. Cataphoric reference, though, is more common in academic prose (Biber, et al., 1999, p. 266).

40% of definite noun phrases are cataphoric reference.

Substitution may be anaphoric or cataphoric. Cataphoric substitution is considerably less frequent than anaphoric, and is usually found across clause boundaries rather than sentence boundaries (Carter & McCarthy, 2006, p. 248).

80% of nouns used in *of*-phrases are inanimate concrete nouns (Biber, et al., 1999, p. 303).

Restrictive clauses are relatively common in academic prose (Biber, et al., 1999, p. 603).

85% of relative clauses in academic prose are restrictive.

Almost 100% of *that* clauses and zero clauses are used with restrictive clauses.

Only 10% of restricted clauses omit the relativizer (Biber, et al., 1999, p. 609-611).

More than 50% of nouns use *who* as a relativizer; less than 25% use zero; less than 5% use *that* ((Biber, et al., 1999, p. 613).

Prepositions

Prepositions are the most common function word class in academic prose (Biber, et al., 1999, p. 92).

80% of noun post modifiers are prepositional phrases, likely relative clauses (Biber, et al., 1999, p. 606).

90% of these prepositions are *of* (65%), *in* (10%), *for*, *on*, *to*, *with* (3-5%)

Embedded prepositional phrases are common (Carter & McCarthy, 2006, p. 270).

of-phrase is used heavily for the genitive in academic prose (Biber, et al., 1999, p. 301-302).
apostrophe *s*-phrase is rarely used in academic prose

Prepositional verbs are relatively common in academic prose, those less so than in other registers (Biber, et al., 1999, p. 415).

of-which and *whose* equally distributed in academic prose (Biber, et al., 1999, p. 618).

Preposition+*which* relativizers are only common in academic prose (Biber, et al., 1999, p. 625).

60% of process adverbs are used in a prepositional phrase (Biber, et al., 1999, p. 787).

50% of circumstantial adverbials are a prepositional phrase, with 80% in final position (Biber, et al., 1999, p. 802).

Structures with a stranded preposition is most common form in all registers except academic prose. Only 20% of prepositions in academic prose are stranded (Biber, et al., 1999, p. 106).

Preposition of contrast *in spite of/despite* (formal), *notwithstanding* very formal (Leech & Svartvik 2002, p. 113).

Adjectives

Greater than 95% of noun pre modifiers in academic prose have the pattern adv+adj+head noun (Biber, et al., 1999, p. 599).

Derived adjectives are much more common in academic prose, especially *-ed* adjectives (Biber, et al., 1999, p. 531).

60% of pre modifiers are adjectives (Biber, et al., 1999, p. 589).

Several pre modifiers often occur, combining both adjective and noun phrase modifiers (Carter & McCarthy, 2006, p. 268).

Quantifiers are used similarly across registers (Biber, et al., 1999, p. 277).

Hyphenated compound adjectives are frequent (Carter & McCarthy, 2006, p. 269).

Adjectives are frequently modified by adverbs in academic style (Carter & McCarthy, 2006, p. 269).

In scientific and technical writing, adjectives of classification and noun phrase pre modifiers are frequent. Several pre modifiers often occur, combining both adjective and noun phrase modifiers (Carter & McCarthy, 2006, p. 268).

In humanities subjects, evaluative adjectives are more frequent where opinion and personal stance are often foregrounded. Such adjectives are normally gradable and may be pre modified by adverbs of degree (Carter & McCarthy, 2006, p. 268).

Greater than 75% of adjectives in academic prose are attributive adjectives (Biber, et al., 1999, p. 506).

Most of these are classifiers, relational and topical.

Most common attributive adjectives in academic prose: *simple, basic, common, following, higher, individual, lower, particular, similar, specific, total, various, local, natural, normal, oral, physical, public, sexual*. *-al* suffix particularly common among attributive adjectives.

-ing and *-ed* adjectives are rare in academic prose.

The comparatively *-er than* form of adjectives is extremely common in academic prose (Biber, et al., 1999, p. 524).

Comparative is used more than in any other register.

Superlative is rare in academic prose.

Fewer with plural count nouns is preferred in formal contexts (Leech & Svartvik 2002, p. 286).

Adverbs

The greatest number of adverbs in academic prose are degree and linking adverbs (Biber, et al., 1999, p. 540).

55% of adverbs in academic prose have *-ly* suffix.

Greater than 30% of adverbs are single adverbs.
good is never used as an adverb in academic prose.

The adverb *approximately* is used primarily in academic prose (Biber, et al., 1999, p. 113).

Academic prose makes use of degree modifiers (whereas conversation prefers amplifiers) (Biber, et al., 1999, p. 565).

Almost 90% of adverbials in academic prose are circumstantial (although other registers use circumstantial adverbials more) (Biber, et al., 1999, p. 76).

50% of these are single adverbs, with 75% used in medial position.

50% are a prepositional phrase, with 80% used in final position.

70% are manner; 80% of these are in final position.

Greater than 60% are 4+ words long (Biber, et al., 1999, p. 808).

Finite clauses are the most common type of circumstantial adverbial in all registers (though they are used less in academic prose than other registers) (Biber, et al., 1999, p. 826).

They appear in final position in academic prose (Biber, et al., 1999, p. 835).

Linking adverbials only account for less than 10% of all adverbials in academic prose, although this is still a greater percentage than in other registers (Biber, et al., 1999, p. 765).

Greater than 60% of linking adverbials are single adverb (Biber, et al., 1999, p. 769).

Greater than 20% are prepositional phrases.

50% of linking adverbials are in initial position with 40% in medial position (Biber, et al., 1999, p. 884).

Greater than 40% of linking adverbials are result/inference; less than 30% apposition.

Linking adjuncts are more frequent in formal styles and in writing (Carter & McCarthy, 2006, p. 290).

The following linking adjuncts are used in academic English and not frequent in conversation: *additionally, in addition, similarly, equally, likewise, furthermore, moreover, accordingly, hence, in view of this/that, as a consequence, in consequence, therefore, as a result, in (the) light of this/that, thus, consequently, by/in contrast, nevertheless, on the other hand, conversely, nonetheless, however, on the contrary, finally, in conclusion, in summary, firstly etc, lastly, in short, respectively, in brief, in sum, subsequently* (Carter & McCarthy, 2006, p. 290).

Single adverbials are used differently in academic prose than other registers (Biber, et al., 1999, p. 846).

65% of *as* used for manner (*as* is the most common subordinator in academic prose, and overall)

95% *since* used for reason

80% *while* used for concession

Non-finite subordinate clauses are particularly common, and ellipted subordinate clauses are also frequent (Carter & McCarthy, 2006, p. 289).

Although is used more than 3 times in academic prose than *though* (Biber, et al., 1999, p. 845).

In academic prose, *though* is used 95% of the time as a subordinator (but 90% of the time as a linking adverbial in conversation) (Biber, et al., 1999, p. 850).

Academic prose makes use of stance less than other registers; stance adverbials are less than 1% in academic prose (compared to 10% in conversation) (Biber, et al., 1999, p. 766).

Sentence adverbials expressing the meaning *in spite of this/that*: *yet, however, nevertheless* (formal), *all the same* (informal) (Leech & Svartvik 2002, p. 114).

In written English, we often use non-finite and verbless clauses as adverbials and modifiers (Leech & Svartvik 2002, p. 15-16).

For a stronger and more emphatic contrast, in formal English we find occasionally a sentence adverbial with coordination or subordination (Leech & Svartvik 2002, p. 190).

In writing and formal speech you can list a series of points by such adverbs as *first/ly, secondly, next, last/ly, finally*. Phrases such as *to begin with, in the [second] place, to conclude* can also be used. Similar to these adverbials are *also, moreover, furthermore, what is more* which indicate that an additional point is being made (Leech & Svartvik 2002, p. 188).

A more formal way to express manner is *in a ...way/manner* or *with*+abstract noun phrase (as opposed to just adverb (*ly*) or adverb phrase) (Leech & Svartvik 2002, p. 105).

on account of, therefore, thus, accordingly, hence, consequently, participle clauses are formal expressions of cause/reason (*so* is informal) (Leech & Svartvik 2002, p. 109-110).

furthermore, what is more are formal to express reinforcement or another point in an argument. Informal *besides, in any case, in fact, anyway* (Leech & Svartvik 2002, p. 188).

Pronouns

Pronouns are rare overall in academic prose (Biber, et al., 1999, p. 92).

Substitute forms include indefinite quantifying pronouns.

Existential *it* is used in academic prose, though even this is rare.

95% of existential *there* clauses in academic prose use *be* (Biber, et al., 1999, p. 945).

In academic prose, *exist* is the alternative to *be* (but other registers don't use this).

Generic *one* is used most frequently in all registers; *one* is used most frequently in academic prose, and more in academic prose than other registers (Biber, et al., 1999, p. 580).

one and *some/ones* are the most common items used to substitute for count nouns, *some* for noncount. *That* and *those* are used as substitutes in formal contexts (instead of *one*) for non-count nouns where the substitute of the noun phrase is taking the place of a definite noun phrase (Carter & McCarthy, 2006, p. 250-251).

Especially in academic style, *that/those of* is used instead of *one*. Is preferred to possessive. Common only in academic prose (Carter & McCarthy, 2006, p. 251).

That is only used as a demonstrative pronoun 10% of the time in academic prose (Biber, et al., 1999, p. 351).

That is only used as a demonstrative determiner 5% of the time in academic prose. 85% of the uses of *that* are “other” (e.g. relativizer).

Which is the most common relativizer in academic prose (Biber, et al., 1999, p. 609-611).

Locative *there* is rare in academic prose; existential *there* is used 10 times more than locative (Biber, et al., 1999, p. 948).

Determiners

Articles are used more in academic prose than other registers; second most common function word in academic prose (Biber, et al., 1999, p. 92).

Demonstrative determiners, especially *this* and *these* are most common in academic prose (Biber, et al., 1999, p. 247).

The is used three times more than *a/an* (Biber, et al., 1999, p. 270).

85% of articles in subject position are *the*

70% of articles in prepositional complement are *the*

55% of articles as object of a preposition are *the*

do so is generally used in more formal contexts for making general reference to a series of actions or events (Carter & McCarthy, 2006, p. 253).

Cataphoric reference to an adjective complement is also possible with *so*, but it is very infrequent and rather formal (Carter & McCarthy, 2006, p. 254).

With *none of* grammatical concord insists that *none* is singular, but notional concord invites a plural verb. A singular verb is typical of written formal style. Same with *either/neither, -body/-one*. They take *they* in informal. In formal, the tendency has been to use *he*, but is now more *s/he* or *he or she* (Leech & Svartvik 2002, p. 276).

In more formal use, *that/those* (but not *this/these*) can function as relative antecedents i.e. the word the pronoun refers to (Leech & Svartvik 2002, p. 280).

Verb Types

Overall, verbs are used less in academic prose than other registers (Biber, et al., 1999, p. 62; 456).

Greater than 50% of *that* complement clauses are post predicate. Patterns include verb+*that* clause, verb+np+*that* clause, verb+ to np+*that* clause (although to np isn't common in academic prose) (Biber, et al., 1999, p. 661).

Lexical verbs are less common in academic prose than other registers, but they are still the most common verb type for academic prose (Biber, et al., 1999, p. 62).

Be is the most used copula in academic prose (as well as overall) (Biber, et al., 1999, p. 359).

The patterns *be+adjective* and *become+adjective* are extremely productive (Biber, et al., 1999, p. 437).

Academic prose uses mental verbs less than other registers (Biber, et al., 1999, p. 367).

Although activity verbs make up 49% of all common verbs, in academic prose existence verbs are used as much as activity verbs (and more so than in conversation) (Biber, et al., 1999, p. 366).

Academic prose reports relations among entities—both concrete and abstract—using simple statements of existence/relationship or occurrence (Biber, et al., 1999, p. 372).

Compared to other registers, academic prose reports relatively few physical, mental, or communication activities. When academic prose does report these activities, they're attributed to inanimate objects (Biber, et al., 1999, p. 372).

Greater than 60% of causative, occurrence, and existence verbs have an inanimate subject (Biber, et al., 1999, p. 378).

~30% activity verbs, 20% communication verbs, and 10% mental verbs have an inanimate subject (Biber, et al., 1999, p. 378).

The verb patterns SV(O)+complement clause followed by SVO_d noun phrase is most common in academic prose (as well as all registers) (Biber, et al., 1999, p. 385-390).

Activity verbs in academic prose tend to describe static situations or relationships (almost becoming existence verbs). Activity verbs can also have a causative or facilitative sense in academic prose (Biber, et al., 1999, p. 380).

Phrasal verbs are relatively rare in academic prose. Academic prose generally shows a greater reliance on derived verbs and more specialized verbs (Biber, et al., 1999, p. 409).

Prepositional verbs are relatively common in academic prose, though less so than other registers (Biber, et al., 1999, p. 415).

33% of prepositional verbs in academic prose are activity verbs

26% existence

19% mental (Biber, et al., 1999, p. 419).

These figures coincide with type of verb use overall.

Common prepositional verbs in academic prose are typically used in the passive and take the double object pattern 2 (Biber, et al., 1999, p. 421).

The most common pattern is *use NP in*

Have as a transitive main verb is least common in academic prose (although it's more common than any single lexical verb) (Biber, et al., 1999, p. 429).

Do is relatively rare in academic prose, especially compared to other registers (probably because it's interrogative and negative, which are rare in academic prose) (Biber, et al., 1999, p. 432-433).

do so is generally used in more formal contexts for making general reference to a series of actions or events (Carter & McCarthy, 2006, p. 253).

Both simple and progressive forms of *will* may be used to refer forward to outline or point to things which are to be found later in the text (anaphoric) (Carter & McCarthy, 2006, p. 275).

Be going to and *'ll* and other contractions are generally avoided in academic writing (Carter & McCarthy, 2006, p. 276).

Tense/Aspect/Voice

Present tense is used greater than 60% of the time in academic prose. Modals are used approximately 15%, and past tense is used approximately 20% (Biber, et al., 1999, p. 456).

Greater than 90% of verbs in academic prose are in the simple aspect (Biber, et al., 1999, p. 461).

5-6% of verbs use the perfect aspect.

Greater than 80% of these use the *present* perfect.

The progressive aspect is almost nonexistent in academic prose.

In general, progressive forms are not frequent in academic texts.

75% of verbs in academic prose are in the active voice. Although only 25% of verbs in academic prose are passive, academic prose uses the passive voice more than any other register (Biber, et al., 1999, p. 476).

Can and *should* are commonly used with passive voice in academic prose (Biber, et al., 1999, p. 499).

Passive voice is common in academic discourse relative to other registers. Verbs especially common with passive voice in academic prose include *BE made, given, taken, used, found, seen, considered, shown* (Carter & McCarthy, 2006, p. 277).

Active voice verbs do occur frequently in academic discourse (Carter & McCarthy, 2006, p. 277).

Post modifiers in noun phrase with short passives is more than two times as common in academic prose than other registers (though it's not all that common in academic prose) (Biber, et al., 1999, p. 938).

Long passives rarely occur in academic prose, but they don't occur in any other register at all, really, so they're unique to academic prose (Biber, et al., 1999, p. 940).

Get-passives are rare in academic writing (Carter & McCarthy, 2006, p. 278).

Commands with *let* + 3rd person subject are formal, often elevated in style (Leech & Svartvik 2002, p. 266).

3 ways to express hypothetical meaning in subclauses (other than past tense): *were* subjunctive, *were to*+infinitive. *Were to* + infinitive to express hypothetical meaning is rather formal as is *should* + infinitive, both suggesting tentative conditions; limited generally to conditional clause.

another type of hypothetical condition clause has no *if* but begins with an operator before the subject (inversion): subjunctive *were* and putative *should* are formal and can be replaced by an *if*-clause (Leech & Svartvik 2002, p. 147).

Subjunctive is more common in written formal English/is quite common in written English (Leech & Svartvik 2002, p. 396-397).

Modals

Greater than 50% of stance is marked by modals in academic prose (Biber, et al., 1999, p. 979).

Modals are used less than 15,000 times per million words (get percentage) in academic prose (Biber, et al., 1999, p. 491).

Academic prose doesn't make use of semi-modals (Biber, et al., 1999, p. 486).

Other

Coordinating conjunctions and correlative coordinators are not especially common in academic prose, but are used more here than other registers (Biber, et al., 1999, p. 85).

Or is less than 1% of conjunctions in academic prose (Biber, et al., 1999, p. 81).

And/or is not especially common in academic prose, but are used more here than other registers.

Coordination tags *and so on, etc* are common only in academic prose (Biber, et al., 1999, p. 116).

There is no and *There is a* are two of the most common three word lexical bundles in academic prose (Biber, et al., 1999, p. 994).

existential *there* is relatively common in academic prose (Biber, et al., 1999, p. 948).

existential *there* is used 10 times more than locative *there*.

there+exist is found in formal academic styles (Carter & McCarthy, 2006, p. 287).

There are greater than 2500 three word lexical bundles in academic prose. They occur greater than 60,000 times per million words (Biber, et al., 1999, p. 993-994).

The average three word lexical bundle occurs 15 times per million words in academic prose.

21% of the words in academic prose occur in lexical bundles (Biber, et al., 1999, p. 995).

Greater than 60% of lexical bundles in academic prose are nominal (rather than clausal) (Biber, et al., 1999, p. 995).

Idiomatrical phrases are more common in academic prose than conversation (though still not so common) (Biber, et al., 1999, p. 1028).

etc is the most common coordination tag in academic prose (Biber, et al., 1999, p. 116).

The suffix *ize/ise* in academic prose is by far the single most productive affix across registers. Greater than 120 different verb lexemes (Biber, et al., 1999, p. 401).

It cleft is relatively common in all registers, but most common in academic prose. *Wh-* clefts are used in other registers, but are virtually non-existent in academic prose (Biber, et al., 1999, p. 961).

Academic prose makes use of stance less than half the amount of stance in conversation (Biber, et al., 1999, p. 979).

I think is less commonly used in academic writing than in everyday language, and expressions such as *in my opinion/I would suggest, argue/it is reasonable* are preferred (Carter & McCarthy, 2006, p. 278).

Rhetorical questions and questions which are immediately answered are frequent in academic contexts (Carter & McCarthy, 2006, p. 288).

In more formal style, *used not to* is preferred (Leech & Svartvik 2002, p. 257).

Explanations can be made in 3 ways. These expressions are typical of written English. *That is/for example*. The Latin abbreviations *i.e., viz, e.g.* are mainly found in formal written texts (Leech & Svartvik 2002, p. 189).

Definite frequency sometimes uses *per* instead of *a/n* e.g. once per day (Leech & Svartvik 2002, p. 92).

Frequency phrases generally have no prepositions, except *on occasion*, which is rather formal (Leech & Svartvik 2002, p. 92).

Distribution of function word class (Biber, et al., 1999, p. 92).

140kpm prepositions
85kpm determiners
50kpm primary auxiliaries
40kpm coordinators and pronouns
10kpm modals and subordinators
5kpm particles

Distribution of lexical word class (Biber, et al, 1999, p. 65)

300kpm nouns	90kpm verbs
100kpm adjectives	20kpm adverbs

APPENDIX 5

LINGUISTIC FEATURES OF SCHOLARLY WRITING BY FUNCTION

Text reference

- *this* is used for immediate textual reference or to signal a new entity or important topic in the text (Biber, et al., p. 349).
- *one* is used for generic rather than specific reference (Biber, et al., p. 580).
- *it* is used to continue reference to an entity which has already been established as a topic in the text (Carter & McCarthy, 2006, p. 246).
- *that* is used to refer to facts, assertions, and other less emphatic entities (Carter & McCarthy, 2006, p. 246).
- Post modifiers of nouns are used most commonly for first mention of a reference; pre modifiers are used for first and subsequent mentions (Biber, et al., p. 585-586).
- Simple nouns are most used for first reference (Biber, et al., p. 585-586).
- 85% of restrictive clauses help identify the reference of the head noun (Biber, et al., p. 603).
- existential *there* focuses on something of interest that will be picked up by later references in the text (Biber, et al., p. 951).
- *do so* is generally used in more formal contexts to make general reference to a series of action or events (Carter & McCarthy, 2006, p. 253).
- Both simple and progressive forms of *will/shall* may be used to refer forward to outline or point to things which are to be found later in the text (anaphoric) (Carter & McCarthy, 2006, p. 275).
- personal subjects and active voice verbs occur frequently in providing textual signpostings (Carter & McCarthy, 2006, p. 278).
- *we* can also occur in textual signposting to orient the reading in some way (Carter & McCarthy, 2006, p. 285).

Likelihood

- Introductory *it* and *that* clauses express degrees of likelihood in formal English (Leech & Svartvik 2002, p. 150).
- *-ly* adverbs involve likelihood of a proposition (Biber, et al., 1999, p. 541).

Cohesion

- referring expression such as pronouns, determiners, and locative adverbs
- parallel structures (Carter & McCarthy, 2006, p. 263).
- across paragraph boundaries, a definite determiner + noun may be preferable (Carter & McCarthy, 2006, p. 264).
- linking adjuncts (Carter & McCarthy, 2006, p. 264).
- Linking adverbials present and support arguments and overtly mark links between ideas (Biber, et al., 1999, p. 875).

Series

- *do so* is generally used in more formal contexts to make general reference to a series of action or events (Carter & McCarthy, 2006, p. 253).
- Existential *there* introduces a series of elements (Biber, et al., 1999, p. 951).

Assertion/Generalization

- *Can* is used to make fairly confident through not absolute assertions (Carter & McCarthy, 2006, p. 280).
- *Would* is frequently used to hedge assertions someone might challenge and is frequently used with *appear/seem* (Carter & McCarthy, 2006, p. 281).
- *that* is used to refer to assertions (Biber, et al., 1999, p. 694).
- Expressions commonly used to make a claim more assertively (boosting): *categorically, indisputably, plainly, certainly, inevitably, undeniably, clearly, irrefutably, undoubtedly, definitely, observably, unquestionably, emphatically, obviously* (Carter & McCarthy, 2006, p. 284).
- *Will* is used to make confident predictions or to assert known or accepted facts; *must* is used to make confident predictions or conclusions (Carter & McCarthy, 2006, p. 279).
- Modals may be used to make propositions less assertive (Carter & McCarthy, 2006, p. 279).
- Unmodified simple forms are often used to more directly and more confidently assert a proposition (Carter & McCarthy, 2006, p. 279).
- *Should* is used to hedge conclusions and predictions, but expresses confidence (Carter & McCarthy, 2006, p. 279).
- Expressions commonly used for hedging: *apparently, generally, roughly, arguably, likely, seemingly, broadly, normally, surely, evidently, partially, typically, frequently, probably, usually, as a rule, in a way, in some respects, broadly speaking, in some cases, more or less, in a sense, in principle, in some senses* (Carter & McCarthy, 2006, p. 282).
- The higher frequency of *many* and *some* in academic prose agrees with the need for expressing guarded generalizations (Biber, et al., 1999, p. 277).
- *-ly* adverbs have to do the generalizability of a proposition (Biber, et al., 1999, p. 541).

Processes/Procedures

- The present perfect progressive may be used to refer to an ongoing process in the discourse up to a given point (Carter & McCarthy, 2006, p. 275).
- Passive voice is used to shift the focus from human agency to actions, processes, etc being described (Carter & McCarthy, 2006, p. 277).
- *-ly* adverbs are used in descriptions of processes (Biber, et al., 1999, p. 541).
- *Help* clauses in academic prose are generally bare infinitive used to describe a process that is facilitated or helped by some other factor (Biber, et al., 1999, p. 736).
- Purpose clauses present procedures (Biber, et al., 1999, p. 824-825).
- Simple past is preferred when referring to procedures used in experiments and studies (Carter & McCarthy, 2006, p. 274).

Reporting

- Passive voice is used to report findings or express logical relations (Biber, et al., 1999, p. 480).
- Extraposed noun complement clauses are used to report attitude or stance not overtly attributed to any person (Biber, et al., 1999, p. 674).
- The present simple is used to report outcome, results or findings, and/or major tenets or central aspects of the work of other academics (Carter & McCarthy, 2006, p. 273).

- Simple past is preferred when citations report experiments (Carter & McCarthy, 2006, p. 274).
- The present perfect is used to imply the continuing validity of earlier findings or practices (Carter & McCarthy, 2006, p. 274).

Stance

- *I* is used in academic discourse to refer to one's stance or conclusions or when contrasting one's own approach with that of others; commonly used in this manner with the following verbs: *accept, assume, suggest, advocate, believe, suppose, agree, consider, suspect, argue, propose, think* (Carter & McCarthy, 2006, p. 284).
- Extraposed noun complement clauses are used to report attitude or stance not overtly attributed to any person (Biber, et al., 1999, p. 674).
- *-ly* adverbs (Biber, et al., 1999, p. 541).

Comparisons

- *As* clauses (without an *it* subject) often occur in the passive to exemplify or compare (Carter & McCarthy, 2006, p. 289).
- A more formal construction to compare equivalent tendencies is an adverbial clause of proportion introduced by *as* followed by *so* in the main clause (Leech & Svartvik 2002, p. 124).
- The high use of comparison may reflect the importance of comparison as a means of understanding and explicating reality (Biber, et al., 1999, p. 529).

Summaries

- Non-finite clauses are used to signal summaries (Biber, et al., 1999, p. 291).
- The present perfect is used to summarize points including conclusions (Carter & McCarthy, 2006, p. 275).

Presenting new information

- 85% of *of-phrases* present new information (Biber, et al., 1999, p. 305).
- Existential *there* presents new information (Biber, et al., 1999, p. 943).

Condition

- Purpose clauses make recommendations for improving conditions (Biber, et al., 1999, p. 824).
- *-ly* adverbs used in descriptions of conditions (Biber, et al., 1999, p. 541).
- The formal conjunction *in case of* is often used to express condition (Leech & Svartvik 2002, p. 111).

Information packaging

- lexical bundles with existential *there* are used for information packaging (Biber, et al., 1999, p. 1024).
- To integrate the maximum amount of information in a noun phrase, prepositional phrases are very frequent (Carter & McCarthy, 2006, p. 270).

Precision

- *it* cleft allows precise statements to be made (Biber, et al., 1999, p. 1963).

- Adverbials in academic prose contribute to making information more precise, like giving details about how things are done and providing descriptions of the results of studies (Biber, et al., 1999, p. 785).
- The more frequent use of *both* and *each* reflects concern with precision (Biber, et al., 1999, p. 277).

Existential there

- brings attention to things that are done (Biber, et al., 1999, p. 951).
- focuses on the fact something does (or doesn't) exist (Biber, et al., 1999, p. 951).

Passive voice is used

- to refer to aspects of scientific methodology and analysis (Biber, et al., 1999, p. 480).
- Due to the focus on relationships among inanimate entities, common prepositional verbs in academic prose are typically used in the passive and take the double object pattern 2, *use NP in* (Biber, et al., 1999, p. 421).
- since it is often felt necessary to shift the focus from human agency to the actions, processes, etc being described, especially in academic writing (Carter & McCarthy, 2006, p. 277).
- is particularly prevalent in abstracts (Carter & McCarthy, 2006, p. 277).

Other

The concern of academic prose with abstract concepts may be the reason for the high number of derived nouns used (Biber, et al., 1999, p. 323).

Activity verbs in academic prose tend to describe static situations or relationships (almost becoming existence verbs) rather than actions or events (Biber, et al., 1999, p. 380).

Have is often used to link an inanimate subject to some abstract quality (Biber, et al., 1999, p. 429).

Most of the amplifiers in academic prose express degree of intensity of a specific characteristic (Biber, et al., 1999, p. 566).

where with relative clauses are used to mark logical rather than physical locations (Biber, et al., 1999, p. 626).

And is used as a coordinated adjective to identify two distinct attributes of a single referent, but opposite attributes of a plural referent (Biber, et al., 1999, p. 601).

Noun + *of* + noun is common to denote groups, parts, and quantities (Carter & McCarthy, 2006, p. 113).

To + noun marks human goals, opportunities, or actions (Biber, et al., 1999, p. 652-53).

Certain and *such* are used only in indefinite noun phrases; the former singles out a specific person(s)/thing(s) the later person(s)/thing(s) of a particular kind (Biber, et al., 1999, p. 281).

Formal ways to express hypothetical meaning include *were to* + infinitive and *should* + infinitive (Leech & Svartvik 2002, p. 147).

The cleft sentence with *it* is particularly useful in written English where intonation can't be used (Leech & Svartvik 2002, p. 217).

We and *you* can refer to the academic community in general; *we* can also be used to create a shared sense of community. *You* is frequently used in this manner, though less than *we* (Carter & McCarthy, 2006, p. 285).

Imperatives are often used to invite the reader to pay attention to something (Carter & McCarthy, 2006, p. 288).

Because of the frequent need for definition and speculation in academic English, post modified and complemented noun phrases are extremely frequent (Carter & McCarthy., 2006, p. 269).

APPENDIX 6

NEGATIVE SALIENT LINGUISTIC FEATURES OF SCHOLARLY WRITING

Full relative clauses as well as *to*-clauses are rare in academic prose (Biber, et al., 1999, p. 637).

Noun complement clauses

Wh- clauses never used in academic prose (Biber, et al., 1999, p. 647).

Omission of *that* is rare in academic prose (Biber, et al., 1999, p. 680).

Apostrophe *s*-phrase is rarely used in academic prose (Biber, et al., 1999, p. 302).

-ing and *-ed* adjectives are rare in academic prose (Biber, et al., 1999, p. 533-534).

The superlative is rare in academic prose (Biber, et al., 1999, p. 524).

Full relative clauses, *to*-clauses, and *-ing* and *-ed* clauses are rare as post-modifiers in academic prose (Biber, et al., 1999, p. 606).

Good is never used as an adverb in academic prose (Biber, et al., 1999, p. 543).

Academic prose makes use of stance less than other registers; stance adverbs are rare in academic prose (Biber, et al., 1999, p. 979).

Although prepositional verbs are common in academic prose, they are less so than other registers (Biber, et al., 1999, p. 415).

Pronouns are rare overall in academic prose (Biber, et al., 1999, p. 65).

Locative *there* is rare in academic prose (Biber, et al., 1999, p. 948).

Negative forms (Biber, et al., 1999, p. 159), imperatives (Biber, et al., 1999, p. 221), and questions (Biber, et al., 1999, p. 211) are rare in academic prose.

Cardinal-digital numbers are not common in academic prose (Biber, et al., 1999, p. 278).

Activity verbs in academic prose tend to describe static situations or relationships (almost becoming existence verbs) rather than actions or events (Biber, et al., 1999, p. 380).

Phrasal verbs are relatively rare in academic prose. Academic prose generally shows a greater reliance on derived verbs and more specialized verbs (Biber, et al., 1999, p. 409).

Do is relatively rare in academic prose, especially compared to other registers (Biber, et al., 1999, p. 432-433).

Compared to other registers, academic prose reports relatively few physical, mental, or communication activities. When academic prose does report these activities, they're attributed to inanimate objects (Biber, et al., 1999, p. 373).

Greater than 60% of causative, occurrence, and existence verbs have an inanimate subject (Biber, et al., 1999, p. 378).

~30% activity verbs, 20% communication verbs, and 10% mental verbs have an inanimate subject (Biber, et al., 1999, p. 378).

Academic prose reports relations among entities—both concrete and abstract—using simple statements of existence/relationship or occurrence (Biber, et al., 1999, p. 372).

Structures with a stranded preposition is most common form in all registers except academic prose. Only 20% of prepositions in academic prose are stranded compared to 80% in conversation, where the other 20% aren't usually able to be stranded anyway! (Biber, et al., 1999, p. 106).

Academic prose doesn't make use of semi-modals (Biber, et al., 1999, p. 486).

The progressive aspect is almost nonexistent in academic prose (Biber, et al., 1999, p. 461).

I think is less commonly used in academic writing than in everyday language (Carter & McCarthy, 2006, p. 278).

Be going to, *'ll*, and other contractions are generally avoided in academic writing (Carter & McCarthy, 2006, p. 276).

Get-passives are rare in academic writing (Carter & McCarthy, 2006, p. 278).

Appendix 7A Summary Worksheet

File Wkst	N.P.	No modifier		Pre-modifier		Post-modifier		Both		<i>the</i>		Proper Noun	
	Raw #	Raw #	%	Raw #	%	Raw #	%	Raw #	%	Raw #	%	Raw #	%
1	110	24	21.82%	33	30.00%	31	28.18%	22	20.00%	40	36.36%	44	40.00%
2	116	20	17.24%	28	24.14%	40	34.48%	28	24.14%	43	37.07%	75	64.66%
3	90	10	11.11%	33	36.67%	16	17.78%	31	34.44%	22	24.44%	12	13.33%
4	100	21	21.00%	26	26.00%	23	23.00%	30	30.00%	28	28.00%	39	39.00%
5	153	29	18.95%	44	28.76%	40	26.14%	40	26.14%	53	34.64%	40	26.14%
TOTAL	569	104	18%	164	29%	150	26%	151	26%	186	33%	210	37%

sum=total 100%
 yes yes
 yes yes
 yes yes
 yes yes
 yes yes

File Wkst	Pre-M Noun		Pre-M Adj		Pre-M Adj <i>and</i> Adj		Post-M <i>of</i>		Post-M <i>ot.</i> Prep		Post-M <i>app.</i>	
	Raw #	%	Raw #	%	Raw #	%	Raw #	%	Raw #	%	Raw #	%
1	6	10.91%	49	89.09%	0	0.00%	10	18.87%	18	33.96%	0	0.00%
2	7	12.50%	49	87.50%	0	0.00%	15	22.06%	22	32.35%	0	0.00%
3	14	21.88%	48	75.00%	2	3.13%	20	42.55%	15	31.91%	2	4.26%
4	7	12.50%	47	83.93%	2	3.57%	15	28.30%	17	32.08%	2	3.77%
5	17	20.24%	67	79.76%	0	0.00%	36	45.00%	31	38.75%	0	0.00%
TOTAL	51	16.19%	260	82.54%	4	1.27%	96	31.89%	103	34.22%	4	1.33%

File Wkst	Post-Mod other		Post-M Rel Cl		Omit Relativizer		which Relativizer		that Relativizer	
	Raw #	%	Raw #	%	Raw #	%	Raw #	%	Raw #	%
1	10	18.87%	15	28.30%	4	26.67%	0	0.00%	8	53.33%
2	19	27.94%	12	17.65%	5	41.67%	0	0.00%	7	58.33%
3	6	12.77%	4	8.51%	0	0.00%	1	25.00%	2	50.00%
4	8	15.09%	11	20.75%	4	36.36%	0	0.00%	2	18.18%
5	7	8.75%	6	7.50%	1	16.67%	0	0.00%	5	83.33%
TOTAL	50	16.61%	48	15.95%	14	29.17%	1	2.08%	24	50.00%

Post-M chk Pre-M Chk
 100% yes
 100% yes
 100% yes
 100% yes
 100% yes
 100% yes

Appendix 7B File Worksheet 1

No Modifier	Pre Modifier	Post-Modifier	Both	Proper Noun
Charisma	immense charisma	xxx ability to	a personal magic of	Randall Patrick McMurphy
xxx men	This character's name	xxx epitome of	special popular loyalty or enthusiasm for	McMurphy
xxx men	all xxx men	Xxx way [that]	xxx ultimate boss on	McMurphy
Buddies	their own decisions	people to	ultimate control over	Nurse Ratched
xxx show	Xxx Big Nurse	boss around	those opinions without	McMurphy
xxx men	all rules	xxx men how	a Christ-like figure to	McMurphy
xxx system	their own opinions	respect for	their ability to	McMurphy
xxx men	McMurphy's charisma	xxx freedom [that]	his ability to	R.P. McMurphy
xxx limits	his courage	a man who	their voices from	McMurphy
xxx men	his ultimate selflessness	a disregard for	McMurphy's charismatic ways of	Nurse Ratched
happiness	a gambling fool	xxx lives of	other emotions that	McMurphy
anger	xxx other men	xxx fun that	these emotions that	McMurphy
xxx battle	xxx other men	xxx strength to	a heroic endeavor in	McMurphy
a man	McMurphy's courage	courage by	a little lost boy who	McMurphy
xxx men	his actions	a battle [that]	such acts of	Robert Boyers
Lady	his actions	Xxx thing [that]	her message on	Harding
Fear	his actions	xxx men that	his originally intendend goal of	Harding that
A savior	his disruptive behaviour	courage by	McMurphy's final act of	Harding
a savior	his hero	an example for	their infamous party on	Xxx Harding at
selflessness	This act	charisma in	xxx two whores from	Nurse Ratched in
Silence	Cheswick's desperate act	people to	a good time with	Harding
xxx men	xxx other men	xxx courage to	xxx other inmates that	Nurse Ratched what
xxx men	A footloose westerner	xxx limits of		McMurphy
xxx respect	their own truths and identities	xxx inmates of		Cheswick to
	his friends	a crush on		McMurphy
	xxx entire ward	xxx voice out		Cheswick
	his mother	courage and strength that		Nurse Ratched (x2)
	this victory	Xxx men on		McMurphy
	her breasts	xxx men that		McMurphy
	this act	xxx importance of		McMurphy
	their voices	xxx men that		McMurphy
	his mind			Billy Bibbit
	a true savior			McMurphy
				Billy with
				Billy
				McMurphy

Appendix 7B File Worksheet 1 (cont'd)

Proper Noun (cont'd)	Pre-M Noun	Pre-M Adj	Pre-M Adj (cont'd)	Pre-M Adj and Adj
Chief Bromden	McMurphy's charisma	immense charisma	his courage	
Chief Bromden	McMurphy's courage	This character's name	his ultimate selflessness	
Nurse Ratched	Cheswick's desperate act	all xxx men	their own truths and identities	
McMurphy	a Christ-like figure to	Xxx Big Nurse	his friends	
McMurphy	McMurphy's charismatic ways of	all rules	their voices	
McMurphy	McMurphy's final act of	a gambling fool	his mind	
McMurphy		xxx other men	her breasts	
		xxx other men	their ability to	
		his actions	his ability to	
		his actions	their voices from	
		his actions	her message on	
		his disruptive behaviour	his originally intendend goal of	
		his hero	their infamous party on	
		This act		
		xxx other men		
		xxx entire ward		
		A footloose westerner		
		this victory		
		this act		
		a true savior		
		a personal magic of		
		special popular loyalty or enthusiasm for		
		xxx ultimate boss on		
		ultimate control over		
		those opinions without		
		other emotions that		
		these emotions that		
		a heroic endeavor in		
		a little lost boy who		
		such acts of		
		xxx two whores from		
		a good time with		
		xxx other inmates that		
		their own decisions		
		their own opinions		
		his mother		

Appendix 7C File Worksheet 2

No Modifier	Pre Modifier	Post-Modifier	Post-Modifier (cont'd)	Both	Proper Noun
a character	many people	a character with	a life that	a new admission to	Jesus Christ
Xxx hospital	many miracles	a savior in	xxx patients how	A major characteristic that	Jesus
xxx hospital	their daily lives	xxx patients at	a way of	his goals in	Jesus
Buddies	many Christ like characteristics	xxx nurse to	an example on	unknown people with	Ken Kesity
people	a gambling fool	Xxx life of		these people that	Kesity
Kindness	a nice guy	xxx time [that]		his best to	McMurphy
People	this kindness	xxx kindness and love of		xxx other members of	McMurphy
Xxx patients	their lives	a characteristic of		a control panel through	Nurse Ratched
xxx hero	this statement	xxx patients with		xxx very same washtub basin that	McMurphy
xxx gesture	control panel	an effort in		Mack's transferral of	McMurphy
an escape	their lives	a sense of		his self free	McMurphy
a man	McMurphy's lesson	a bit of		xxx biggest impact on	Xxx Bible
People	Mack's role	xxx state [that]		a large guy on	Xxx Bible
xxx patients	This statement	xxx men in		This ironic illustration of	Christ
a woman	a big man	xxx way [that]		his size through	McMurphy
Xxx nurse	xxx outer appearance	xxx patients that		This reaction from	Mac
xxx nurse	his size	xxx men how		xxx other members of	Chief Bromden from
emotions	his promise	xxx object through		xxx other members of	McMurphy
xxx nurse	his time	xxx men that		xxx other patients to	R.P. McMurphy
xxx patients	his help	individuals by		xxx worst thing [that]	McMurphy
	a new uniform	a man by		his life for	McMurphy
	a normal woman	a way of		his life on	Xxx Bible
	her uniform	a need for		this event to	Christ
	any bad act	xxx source of		xxx other patients on	McMurphy
	your destiny	xxx man [that]		xxx nurse's cause so	McMurphy
	all xxx patients	xxx fact that		their selves from	McMurphy
	his mission	xxx uniform to		his actions in	McMurphy
	xxx other patients	an opportunity to		Mac's kindness to	McMurphy
		An article about			McMurphy
		an example of			McMurphy
		an excuse to			Chief
		a lesson so			Chief
		xxx patients how			Jesus
		a way to			Christ
		xxx members of			McMurphy
		xxx patients to			Chief

Appendix 7C File Worksheet 2 (cont'd)

Proper Noun (cont'd)	Proper Noun (cont'd)	Pre-M Noun	Pre-M Adj	Pre-M Adj (cont'd)	Pre-M Adj and Adj
Billy Bibbit	McMurphy	control panel	many miracles	his time	
McMurphy	Christ	McMurphy's lesson	many Christ like characteristics	his help	
Chief	McMurphy	Mack's role	a gambling fool	her uniform	
McMurphy		a control panel through	a nice guy	your destiny	
McMurphy		Mack's transferral of	this kindness	his mission	
McMurphy		xxx nurse's cause so	many people	his goals in	
Chief		Mac's kindness to	this statement	his best to	
McMurphy			This statement	his self free	
McMurphy			a big man	his size through	
McMurphy			xxx outer appearance	his life for	
Chief			a new uniform	his life on	
McMurphy			a normal woman	their selves from	
McMurphy			any bad act	his actions in	
Nurse Ratched after			xxx other members of		
Kesey			all xxx patients		
McMurphy			xxx other patients		
Christ			a new admission to		
McMurphy			A major characteristic that		
Christ			unknown people with		
Chief			these people that		
McMurphy			xxx very same washtub basin that		
McMurphy			xxx biggest impact on		
McMurphy			a large guy on		
Chief			This ironic illustration of		
McMurphy			this event to		
Christ			xxx other patients on		
McMurphy			This reaction from		
McMurphy			xxx other members of		
Scanlon			xxx other members of		
xxx Chief			xxx other patients to		
Mac			xxx worst thing [that]		
Mac			their daily lives		
Christ			their lives		
Christ			their lives		
McMurphy			his size		
McMurphy			his promise		

Appendix 7D File Worksheet 3

No Modifier	Pre Modifier	Post-Modifier	Both
Action	This bold statement	an essence of	Psycho-spiritual healer, Gabrielle Roth
readers	Aha! Moment	xxx groundwork for	this idea of
a defect	Xxx Big Nurse	asylum from	xxx only character who
Sanity	xxx Psychiatric Ward	xxx instigator, manipulat	Other patients on
Caginess	her plan	shadows of	our sense of
xxx world	Nurse Ratched's authority	xxx importance of	his own perception of
a man	A paranoid-schizophrenic	responsibility for	xxx healthiest ways to
Realization	an unreliable narrator	Xxx act of	comfort and false safety in
xxx world	his delusions	Xxx dichotomy of	his own role in
a power	This seclusion	xxx world that	his connection to
	Individual sanity	a form of	Kesey's tools to
	his individual sanity	solace for	xxx thin veil between
	Xxx coping skills and avoidance tactics	a society that	xxx two very objective and interwoven states of
	genuine kindness	a power which	xxx fatally rebellious nature of
	hidden wires	xxx best for	a blocked understanding of
	fair game	a basis for	His poverty of
	xxx hearing-capable, competent Indian		his inability to
	his cover		his perception of
	Chief Bromden's aberrations		baby steps to
	Ratched's way		his sense of
	Bromden's growth		a more humanized existence around
	his environment		any idea of
	This newly fostered responsibility		his absolutely emasculated mentor, McMurphy
	Xxx self-realization		Kesey's exploration of
	xxx control panel		an interweaving, fluctuating ribbon of
	Self-realization		Bromden's declaration of
	his past		xxx control panel out
	his present		xxx undefined nature of
	his future		another breakdown without
	individual sanity		Chief's growth in
	his Native American heritage		his great escape toward
	Chief Bromden's future		
	no definite black outlines		

Appendix 7D File Worksheet 3 (cont'd)

Proper Noun	Pre-M Noun	Pre-M Adj	Pre-M Adj (cont'd)
Kesey	Nurse Ratched's authority	This bold statement	his Native American heritage
Bromden	Individual sanity	Aha! Moment	our sense of
Kesey	Chief Bromden's aberrations	Xxx Big Nurse	his own perception of
Bromden	Ratched's way	xxx Psychiatric Ward	his own role in
Bromden	Bromden's growth	A paranoid-schizophrenic	his future
Bromden	xxx control panel	an unreliable narrator	his connection to
Nurse Ratched for	individual sanity	This seclusion	His poverty of
Bromden	Chief Bromden's future	genuine kindness	his inability to
Bromden	Kesey's tools to	hidden wires	his perception of
Chief Bromden	baby steps to	fair game	his sense of
Bromden	Kesey's exploration of	xxx hearing-capable, competent Indian	his absolutely emasculated mentor, McMurphy
Chief Bromden in	Bromden's declaration of	This newly fostered responsibility	his great escape toward
	xxx control panel out	Xxx self-realization	
	Chief's growth in	Psycho-spiritual healer, Gabrielle Roth	
		Self-realization	
		this idea of	
		xxx only character who	
		Other patients on	
		xxx healthiest ways to	
		xxx thin veil between	
		no definite black outlines	
		xxx two very objective and interwoven states of	
		xxx fatally rebellious nature of	
		a blocked understanding of	
		a more humanized existence around	
		any idea of	
		xxx undefined nature of	
		another breakdown without	
		an interweaving, fluctuating ribbon of	
		her plan	
		his delusions	
		his individual sanity	
		his cover	
		his environment	
		his past	
		his present	

Appendix 7E File Worksheet 4 (cont'd)

Proper Noun	Pre-M Noun	Pre-M Adj
One Flew Over xxx Cuckoo's Nest	Nurse Ratched's friend	Xxx only women
Chief Bromden	Nurse Ratched's actions	mentally ill
Nurse Ratched	Chief Broom's views	Xxx mental institution
Chief Broom	xxx male species	all women
Nurse Ratched	xxx television viewing schedule	This opinion
McMurphy	McMurphy's female friends, two prostitutes	mere objects
McMurphy	a majority vote to	xxx fairer sex
Nurse Ratched		these same humans
Ken Kesey		Other writers
Nurse Ratched		only male patients
Nurse Ratched		a majority vote
Chief Broom		these objects
Nurse Ratched		xxx masculine psyche
Kesey		xxx coldest hearted character
Porter		little room
Porter		an errant child
Kesey		an almost catatonic, half-breed Indian who
Nurse Ratched		a cold, domineering woman who
Kesey		xxx other men on
McMurphy		all situations to
Nurse Ratched		a controlling woman who
Nurse Ratched		a spineless doctor on
Kesey		other opportunities to
Robert Forrey		a bit sexist in
Forrey		any signs of
Nurse Ratched		no attempt to
Nurse Ratched		a good job of
Nurse Ratched		any kinship or kindness toward
Nurse Ratched		all xxx men on
Kesey		an evil bitch out
Kesey		a lowbrow approach to
Kesey		role model for
Kesey		two prostitutes who
Kesey		an ugly reality [that]
Kesey (x3)		her job
Nurse Ratched		her staff

Appendix 7F File Worksheet 5

No Modifier	Pre Modifier	Pre-Modifier (cont'd)	Post-Modifier	Pre-Modifier (cont'd)
a religion	Taoist, or Daoist, Philosophy	one's strength	a set of	a battleground of
a philosophy	its roots	Nurse Ratched's cold, sadistic, calculating attitude	a way of	correlations between
xxx reader	This struggle	McMurphy's strength	xxx basis for	an insight into
xxx point	an overtly feminine physique	McMurphy's vigorous, heated, flamboyant attitude	xxx conflict between	a novel for
doctors	an astute observation	xxx entire book	a group of	a form of
a term	Nurse Ratched's breast size	Most people	xxx characteristics of	
man	Xxx "too-big" breast size	these goals	a time during	
xxx skin	a female superpower	an even richer, deeper meaning	a way of	
tongue	any enemy	These principles	xxx doctors that	
heat	my veins		Xxx reference to	
motion	a very chilly figure		imagery of	
sweat	a cold force		beads of	
work	xxx hot, heavy work boots		xxx characteristics of	
xxx smell	This vivid quotation		xxx epitome of	
xxx Nurse	outward centrifugal force		A man of	
Xxx glass	his point		a man of	
xxx nurse	her face		xxx cage that	
Xxx point	a literal and figurative barrier		a barrier in	
xxx enemy	Xxx glass barrier		xxx barrier that	
a wave	a peaceful protest		xxx men in	
xxx opponents	her own self- discipline		xxx Nurses from	
a custom	McMurphy's goal		Xxx power of	
xxx acutes	her cool		a keystone to	
a revolt	Ratched's composure		xxx fights at	
Muscle	Her face		control of	
Misfortune	terrible, cold face		manipulation to	
xxx ward	Nurse Ratched's seemingly unshakable confidence		xxx men in	
xxx ward	her authority		xxx assumption that	
xxx characters	an unnecessary ward policy		xxx staff with	
	xxx prison system		xxx situation to	
	a much more important plan		a way to	
	Xxx massive insubordination		Xxx cvcle of	
	a bad thing		a target for	
	a frontal lobotomy		xxx aids from	
	xxx other men		xxx effects of	

Appendix 7F File Worksheet 5 (cont'd)

Both	Both (cont'd)	Proper Noun	Proper Noun (cont'd)
an eastern school of	his spirit in	Taoism	Ratched
xxx unknown author of	a magnificent example of	Sun Tzu, xxx author of	a martyr to
xxx hot, active, male energy of	This yang element at	Taoism to	Billy with
xxx cold, calm, female energy of	xxx ancient principles of	Taoism	Kesey
a striking similarity between	Kesey's novel with	Xxx Tao of	McMurphy
eastern mysticism with		Nurse Ratched	
a basic understanding of		McMurphy	
two opposing forces in		Kesey	
xxx fundamental elements of		Nurse Ratched	
a maternal role for		Chief Bromden	
a firm position of		Nurse Ratched	
these doctors with		Nurse Ratched	
no definitive way to		Chief Bromden	
another example of		Yang	
xxx gritty, dry layer of		McMurphy	
xxx man smell of		Chief Bromden	
xxx second day [that]		McMurphy before	
a brilliant example of		Nurse Ratched	
Ratched's grip on		McMurphy	
xxx correct course of		McMurphy	
his hand through		McMurphy	
xxx Nurses Station from		McMurphy	
xxx one place that		Chief Bromden	
a bit of freedom from		Nurse Ratched	
Nurse Ratched's sense of		Nurse Ratched	
Nurse Ratched's control over		McMurphy	
a great source of		McMurphy	
her totalitarian control over		McMurphy	
her jealously guarded "cool" within		Xxx I Ching	
another man at		Nurse Ratched's Yin	
a wonderful example of		Nurse Ratched	
Nurse Ratched's calm sense of		Billy Bibbit into	
a quite sane man throughout		McMurphy	
Xxx other men of		McMurphy	
an even more aggressive or offensive way to		McMurphy	

Appendix 7F File Worksheet 5 (cont'd)

Pre-M Noun	Pre-M Adj	Pre-M Adj (cont'd)	Pre-M Adj and Adj
Taoist, or Daoist, Philosophy	This struggle	these doctors with	
Nurse Ratched's breast size	an overtly feminine physique	a firm position of	
Ratched's composure	an astute observation	no definitive way to	
McMurphy's goal	Xxx "too-big" breast size	another example of	
Xxx glass barrier	a female superpower	xxx gritty, dry layer of	
Nurse Ratched's seemingly unshakable confidence	any enemy	a great source of	
xxx prison system	a very chilly figure	xxx second day [that]	
McMurphy's strength	a cold force	a brilliant example of	
Nurse Ratched's cold, sadistic, calculating attitude	xxx hot, heavy work boots	a bit of freedom from	
McMurphy's vigorous, heated, flamboyant attitude	This vivid quotation	xxx correct course of	
xxx man smell of	outward centrifugal force	another man at	
Ratched's grip on	terrible, cold face	a wonderful example of	
xxx Nurses Station from	a literal and figurative barrier	xxx one place that	
Nurse Ratched's sense of	a peaceful protest	a quite sane man throughout	
Nurse Ratched's control over	an unnecessary ward policy	Xxx other men of	
Nurse Ratched's calm sense of	xxx entire book	an even more aggressive or offensive way to	
Kesey's novel with	Most people	This yang element at	
	these goals	a magnificent example of	
	an even richer, deeper meaning	xxx ancient principles of	
	These principles	its roots	
	an eastern school of	my veins	
	xxx unknown author of	his point	
	xxx hot, active, male energy of	her face	
	xxx cold, calm, female energy of	her own self-discipline	
	a striking similarity between	Her face	
	eastern mysticism with	her cool	
	a basic understanding of	her authority	
	two opposing forces in	his hand through	
	a much more important plan	her totalitarian control over	
	Xxx massive insubordination	her jealously guarded "cool" within	
	a bad thing	his spirit in	
	a frontal lobotomy		
	xxx other men		
	one's strength		
	xxx fundamental elements of		
	a maternal role for		

Appendix 7F File Worksheet 5 (cont'd)

Post-M of	Post-M ot. Prep	Post-M app.	Post-Mod other	Post-M Rel Cl
a set of	xxx basis for		manipulation to	xxx doctors that
a way of	xxx conflict between		a keystone to	xxx cage that
a group of	xxx men in		Xxx reference to	xxx barrier that
xxx characteristics of	xxx Nurses from		xxx situation to	xxx assumption that
a way of	xxx fights at		a way to	xxx second day [that]
imagery of	a time during		no definitive way to	xxx one place that
beads of	a barrier in		an even more aggressive or offensive way to	
xxx characteristics of	xxx men in			
xxx epitome of	xxx staff with			
A man of	a target for			
a man of	xxx aids from			
Xxx power of	an insight into			
control of	correlations between			
a firm position of	a novel for			
a form of	a striking similarity between			
an eastern school of	eastern mysticism with			
xxx unknown author of	two opposing forces in			
xxx hot, active, male energy of	a maternal role for			
xxx cold, calm, female energy of	his hand through			
a basic understanding of	these doctors with			
xxx fundamental elements of	Ratched's grip on			
xxx effects of	xxx Nurses Station from			
a battleground of	a bit of freedom from			
Xxx cycle of	Nurse Ratched's control over			
a magnificent example of	her totalitarian control over			
xxx ancient principles of	her jealously guarded "cool" within			
Xxx other men of	another man at			
a wonderful example of	a quite sane man throughout			
Nurse Ratched's calm sense of	This yang element at			
Nurse Ratched's sense of	Kesey's novel with			
a great source of	his spirit in			
another example of				
xxx gritty, drv layer of				
xxx man smell of				
xxx correct course of				
a brilliant example of				

