ENHANCING THE ACQUISITION OF DISCIPLINE-SPECIFIC VOCABULARY

THROUGH STUDENT CONCORDANCING

By

David Longstreth

A Project Presented to

The Faculty of Humboldt State University

In Partial Fulfillment of the Requirements for the Degree

Master of Arts in English: Composition Studies and Pedagogy

Committee Membership

Dr. Suzanne Scott, Committee Chair

Dr. Joseph Diémé, Committee Member

Dr. Janet Winston, Graduate Program Coordinator

December 2017

ABSTRACT

ENHANCING THE ACQUISITION OF DISCIPLINE-SPECIFIC VOCABULARY
THROUGH STUDENT CONCORDANCING

David Longstreth

This Master's Project focuses on advanced-level English learners' acquisition of discipline-specific vocabulary as they transition from intensive English programs into English-medium university coursework. During this period, the number of discipline-specific terms students must master quickly and independently can be overwhelming. To address this problem, this M.A. Project argues that vocabulary-acquisition strategies should be foregrounded in intensive English programs, and that instructors should train students to supplement traditional vocabulary learning methods with independent concordancing strategies. Using concordancers, students can research vocabulary items by scanning a corpus (a large collection of texts) to retrieve examples of discipline-specific terms within authentic texts, revealing patterns of usage and collocation, and facilitating deeper knowledge of new lexical items that can result in more accurate production.

Although many applied linguists have promoted student concordancing, few teaching resources are available on the topic. Therefore, this project outlines an instructional unit scaffolding the process of independent student concordancing. It provides criteria for teachers to consider when selecting a corpus to suit instructional

contexts and aims. It provides an overview of the Corpus of Contemporary American English, a large corpus that is freely accessible online, and it examines the features of its integrated concordancer that can help students learn to utilize corpus data for vocabulary learning. Finally, the project relates the writer's tentative steps in introducing students to concordance data in his teaching, and it presents his experience using corpus-based tools in his own second-language academic writing.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 From Intensive English Program to English-Medium University: A Lexical Leap

For an advanced-level international student, leaving an Intensive English Program and matriculating into English-medium university courses can be as intimidating as it is exhilarating. It is a moment that represents a step into the unknown, no matter how well-prepared or motivated the student is. A learner at this critical stage must demonstrate Cognitive Academic Language Proficiency, or CALP (Cummins 1979), in reading, writing, listening, and speaking, and, on the lexical level, should demonstrate increasing control of a body of terms commonly found across a range of academic disciplines. In other words, students will need to utilize the lexicon of the Academic Word List (AWL) (Coxhead 2000) or the Academic Vocabulary List (AVL) (Gardner and Davies 2014), two valuable resources for learning the language of academia (see Appendix A for a list of abbreviations used throughout this project).

A student beginning to specialize in a particular academic discipline must also build a mental lexicon of its key terms—often categorized as *technical* or *discipline-specific* vocabulary—and these generally consist of specific senses of AWL/AVL terms as well as other specialized or technical words that occur with greater frequency within specific disciplines than they do in more general contexts (Coxhead and Nation 260). For example, the term "stratified" is likely to be used in two different discipline-specific

senses in geology and sociology, but might occur only incidentally—if at all—in mathematics.

The process of discipline-specific vocabulary acquisition can be challenging for language learners as they enter the university because vocabulary acquisition at this stage is primarily based on independent learning strategies; in other words, the student can no longer depend on Intensive English Program (IEP) instructors to provide lists of terms to study. Furthermore, to move beyond mere comprehension of key discipline-specific terms and to begin using them accurately in speaking and writing, a learner needs deep lexical knowledge that comes from repeated exposure to these items in authentic contexts (Nation 1990, 2001) and, I would argue, from strategic independent investigation of— and practice with—newly encountered vocabulary.

One strategy that shows promise in enhancing students' vocabulary acquisition is the use of *concordancers* to research unfamiliar terms (Johns 1991; Flowerdew 1996; Cobb and Horst 2001; Yoon and Hirvela 2004; Gavioli 2005; O'Sullivan and Chambers 2006; Tongpoon 2009; Boulton 2010; Reppen 2010). *Concordancers* are digital tools which allow users to view key words in authentic language contexts (see Figure 1 below). Dee Gardner, an applied linguist specializing in vocabulary acquisition and corpus linguistics, explains that a concordancer is "an electronic tool that allows users to bring together multiple sentences containing the same word or phrase from a text or corpus" (Gardner 189). More specifically, applied linguist Maggie Sokolik defines a concordancer as a "type of index that searches for occurrences of words, parts of words, punctuation, affixes, phrases, or structures within a corpus and can show the immediate

context of the search term" (417). Thus, concordancers allow linguists, researchers, teachers, and students to locate specific language items and see how they behave in their natural environments—in authentic texts contained in a corpus. *Concordancing*, for our purposes, is simply the use of a concordancer to research language phenomena.

Figure 1: Screenshot of Concordance Lines for the Term "Stratified" (WordAndPhrase)

A *corpus*, according to Gardner, is "a collection of materials (written or spoken) compiled for the purpose of linguistic investigation" (189). Randi Reppen, an applied linguist who has written extensively on corpus linguistics, vocabulary, and teacher training, defines it as a "large, principled collection of naturally occurring texts (spoken or written) stored electronically" (2). Corpora have been compiled from texts in many of the world's languages, and cover many varieties and genres of spoken and written communication. Thus, corpora range in size from tiny corpora containing, for example, writing samples collected from students in a single class for the purpose of error analysis

to the frequently updated Corpus of Contemporary American English (COCA) which presently consists of over 560 million words (COCA).

Corpora and concordancers allow users to examine target vocabulary items in authentic contexts, and one benefit of this technology is that it provides access to information about vocabulary usage. More specifically, this technology can reveal how target words are used in conventional patterns of *collocation*. According to Gardner, *collocation* is "the condition that exists when two or more words consistently co-occur…within a certain distance of each other in actual linguistic contexts" (189). In other words, target vocabulary words are likely to be found embedded in particular patterns of *collocates*, or specific words which precede and follow them. For example, for the word "vast," the collocate which occurs most frequently in the COCA is "majority," as in the phrase, "the vast majority." The prevalence of this multiword item highlights the value of collocation in vocabulary learning; knowledge of collocation allows learners to produce terms in ways that feel natural or correct to experienced speakers of a language.

I argue that instructors in IEPs should train their students to use concordancers to supplement traditional independent vocabulary-learning strategies such as using context to infer meaning, analyzing roots and affixes, looking up terms in dictionaries and glossaries, and keeping a lexical journal. Moreover, I argue that concordancing can aid students in acquiring the depth of lexical knowledge necessary to move terms from the *receptive* lexicon—the words a learner comprehends when they are encountered in reading and listening—to the *productive* lexicon, or the words the student can accurately utilize in speech and writing.

To illustrate language learners' difficulties in acquiring discipline-specific vocabulary and explain how a real-world teaching problem led me to investigate corpora and concordancers as a possible solution, it may be helpful to provide an anecdote from my own teaching experience.

I worked for several years as an instructor of English for Academic Purposes (EAP) in an IEP for international students at a university in California. There, I once had a student, "Ahmed" (a pseudonym), who, after matriculating into English-medium university coursework, commented on the lexical challenges of entering an American graduate program, thus inspiring me to pursue this line of research. I had been Ahmed's grammar and writing instructor for several semesters as he progressed through the program's proficiency levels, and I also worked with him in a reading and vocabulary class and in an elective course called "Advanced Vocabulary." Ahmed was consistently a highly motivated student with strong metalinguistic knowledge of language learning, and he consistently went beyond our coursework and studied independently to prepare for the demanding Masters' program in sociology in which he intended to enroll. In our Advanced Vocabulary course, he had eagerly engaged in studying the Academic Word List and practiced independent vocabulary-learning strategies. Each week when I checked my students' lexical journals documenting the terms they had found in texts from the fields in which they aimed to study, his entries were always the most detailed. In many ways, Ahmed was an ideal student who seemed well-prepared to begin studying in the university. However, even a student like Ahmed faces lexical hurdles (Corson 1985) when transitioning into English-medium university studies.

When I checked in informally about how his first two months of graduate school had been going, Ahmed explained that he felt that he was adapting well overall, but that he was having some trouble with all the new vocabulary he was encountering. He explained that he was continuing to carefully learn and document new terms in his discipline as he encountered them, so he had few problems *comprehending* these terms when he came across them again in his readings and lectures. He said, however, that he was becoming frustrated with his ability to *utilize* these new words to demonstrate his competence when speaking and writing. He offered an explanation along the lines of, "I spend all day reading, and every new word I look up, translate, put in a journal, watch YouTube videos about it, but when I need to use the word in class or in writing, I can't remember how to use it." He asked me what he could do to solve the problem.

I wanted to provide an answer.

I was sure of one simple, but not-terribly-helpful, answer: repeated exposure to the target words. Ahmed's time was limited, however, and there was no guarantee that he would incidentally happen upon key terms in context enough times to provide the depth of knowledge needed to produce them when the time came to demonstrate his understanding of the material covered in his courses (Gardner 124). I began thinking about ways he could effectively seek meaningful exposure to target vocabulary words in authentic contexts. It seemed that concordancing could be a viable solution, and I turned to the research literature to see whether there is any evidence to support this approach.

As it turns out, Ahmed was not alone in facing a lexical hurdle as he transitioned into university coursework in English; English learners (ELs) worldwide face similar

challenges as they enter English-medium universities (Cobb and Horst 2001). In fact, research shows that lexical errors are among the most frequent error types found in non-native speaker (NNS) student writing. For example, Dana Ferris (2003), an expert in second-language writing, analyzed errors in texts written by 92 university English as a Second Language (ESL) composition students in the US and found that word-choice errors were the second-most-prevalent error type, accounting for 11.5% of total errors identified (148). Unfortunately, lexical errors in NNS student writing can mark a student as an outsider to a particular academic discipline. According to genre and writing scholar John Swales (1990), a specific lexis is one of the six key characteristics of any discourse community (26). Therefore, a NNS seeking entry into the discourse community of an academic discipline would be well-advised to devote serious attention to mastering its vocabulary as efficiently and effectively as possible.

Furthermore, lexical errors can interfere greatly with the comprehensibility of the writing of NNS university students (Santos 1988). Terry Santos (1988) reported on 178 professors' evaluations of two texts written by NNS university students and found that the professors deemed lexical errors the most serious of all the errors in the students' papers, contributing to their overall impression that the texts were "academically unacceptable" (69). Clearly, bringing academic and discipline-specific vocabulary into productive use as fast as possible should be a learning priority for Ahmed and other EL students worldwide.

Ahmed's comments also reveal that he intuitively grasped another central aspect of learning new words: the difference between *receptive* and *productive* vocabulary. The

number of words a learner comprehends in listening and reading is always going to be higher than the number of words that the student can produce in speaking and writing (Lee and Muncie 297). According to Laufer and Goldstein (2004), two scholars who have researched extensively in the area of vocabulary acquisition, a student's "strength of knowledge" of any word can be conceived of as a hierarchy with four skill levels, which are ranked below from *most difficult* to *least difficult*:

> (a) active recall, or the ability to supply the target word
>
> (b) passive recall, or the ability to supply the meaning of a target word
>
> (c) active recognition, or the ability to recognize the target word when given its meaning, and
>
> d) passive recognition, or the ability to recognize the meaning of a target word when given meaning options (406-407)

For language learners entering the discourse communities of their chosen disciplines, the key terms used within those disciplines must be brought within a student's range of active-recall vocabulary. With that in mind, an important goal for EAP instructors is to determine the best ways to "bring learners' vocabulary knowledge into communicative use" (Laufer and Nation 308). I argue that independent student concordancing is a strategy that can facilitate this process.

1.2 Statement of Purpose

In order to promote deeper and more productive knowledge of discipline-specific lexical items, this M.A. Project will outline a strategy and method for students to

supplement traditional methods of vocabulary acquisition with independent concordance-based research. I argue that students can learn to use concordancing applications which allow them to research unfamiliar vocabulary items and retrieve examples of the ways the words are authentically used in written texts in their chosen disciplines. This approach to vocabulary is designed specifically to benefit advanced-level ELs as they acquire discipline-specific academic vocabulary while transitioning from intensive English programs to English-medium university coursework. I focus on this population of students because they are likely to have highly developed language-learning skills and strategies along with the motivation to become better independent learners based on their impending launch into English-medium university studies. Moreover, these students will soon need to acquire a great deal of specialized vocabulary in their fields.

<div align="center">1.3 Overview of the Project</div>

My approach to this problem draws from research within the fields of corpus linguistics, second-language acquisition, and the teaching of English as an additional language and English for Academic Purposes (EAP).

I begin Chapter Two, the Review of Literature, with a section entitled "Foregrounding Vocabulary in English for Academic Purposes," in which I synthesize scholarly sources to establish the importance of effective vocabulary-acquisition strategies for ELs beginning their university careers. I draw from existing research to outline the principles of academic vocabulary knowledge and to illustrate the challenges learners face in meeting the lexical demands of specialized academic discourse

communities, particularly as they learn to employ academic vocabulary in productive use. In the second section of the Review of Literature, "Principles for Teaching Lexical Issues in English for Academic Purposes," I provide a brief overview of advanced IEP vocabulary instruction and explain how teachers can enhance this process using insights from corpus linguistics. I focus on the kinds of rich information that corpora and concordancing can provide learners by presenting words in authentic contexts, in which words' collocational behavior (i.e., the way these words tend to be found within patterns of other, specific words) can be observed, allowing for inferential learning and discovery of patterns in natural language.

In Chapter Three, "Practical Applications: Preparing Advanced Language Learners to Investigate Vocabulary through Concordancing Technology," I suggest aims and offer an instructional approach for a teaching unit on independent student concordancing. I outline a framework for teachers to consider when choosing a corpus to use with their students. I then provide a system for instructing language learners in using the COCA, a free and readily available corpus and suite of concordancing tools. The goal of this practical section is to help teachers train their students to become independent vocabulary investigators capable of researching key terms within their disciplines in order to use them more accurately in their own writing.

In Chapter Four, "Experiences with Concordancing as a Language Teacher and Language Learner," I present my experience with corpora and concordancing from the dual perspective of someone who is simultaneously a language teacher and a language learner. In the *language-teacher* section, I reflect on the successes and challenges that

accompanied my first steps in integrating concordancing into the curriculum of an

advanced level ESL vocabulary course. In the *language-learner* section, I discuss my

tentative steps in using online search tools and corpus-based web applications as

resources for improving my vocabulary use in academic writing in my second language

(L2), French.

Finally, in Chapter Five, the Conclusion, I review the key points of my argument,

review some guidelines for implementing independent student concordancing, and

discuss possible areas for future research and development of practical resources for

teachers to draw from in the future.

1.4 Significance of the Project

English is the lingua franca of academia internationally, and the numbers of EL

students around the world seeking proficiency in academic English in order to pursue

their goals continue to rise. Therefore, teachers need greater understanding of the ways

students can utilize the immense and ever-expanding amount of natural-language data

digitally available through corpora. Although computerized analyses of text have

influenced linguistics for decades (Biber 1988), and linguists and second-language-

acquisition specialists have argued for the potential value of student concordancing to

facilitate the language-learning process since the 1980s (Boulton 2010), only a few

practical teacher- and student-friendly resources incorporating corpora and

concordancing into the curriculum of IEPs exist (namely, Tribble and Jones 1997,

Gavioli 2005, and Reppen 2010). My synthesis of the research on student concordancing

culminates in an outline of a unit ESL and English as a Foreign Language teachers can implement to guide language learners in using concordancing as an independent vocabulary-acquisition strategy.

1.5 Chapter Summary

In this introduction, I have identified a problem faced by English language learners as they begin English-medium university studies: the challenge of acquiring a daunting number of academic and discipline-specific vocabulary items in a short period of time. These students need to comprehend these terms when they encounter them in reading and listening, but they must also learn to produce them accurately in writing and speaking in order to establish themselves as rhetorically credible members of their chosen academic discourse communities. I have hypothesized that, when combined with traditional methods of vocabulary acquisition, independent concordancing strategies can aid students in this process by providing multiple, targeted examples of lexical items being used in authentic contexts. I have provided a brief overview of the subsequent sections of this project and outlined the issues to be discussed. It is my sincere hope that the system outlined in this M.A. Project, a system enabling IEP instructors to guide advanced-level students in using concordancing tools, will benefit real students like my former student, Ahmed.

CHAPTER 2: REVIEW OF LITERATURE

Instructors in English-for-Academic-Purposes (EAP) programs face the challenging task of preparing students to read, write, listen, and speak at advanced levels in a second—or sometimes third, or fourth—language. This preparation includes developing knowledge of a wide range of aspects of academic writing, including discourse structure, grammatical conventions, research strategies, the revision process, and properly incorporating ideas and material from sources. Applied linguists and Teaching English as a Second Language (TESL) scholars have increasingly emphasized the connection between language learners' lexical knowledge and successful writing at the university level (Coxhead and Byrd 129). Although it would seem to be common sense that a solid foundation in academic vocabulary is critical to a student's ability to generate effective writing, the vocabulary-writing connection has not always been foregrounded in discussions of either ESL writing or ESL vocabulary instruction (Lowry 7-8). This section of this M.A. Project provides a summary of some of the most salient scholarly work on the necessity of foregrounding vocabulary skills in EAPs, and it presents a framework of vocabulary acquisition for advanced learners preparing to matriculate into English-medium university courses. This framework comprises some of the metalinguistic knowledge and vocabulary-acquisition strategies necessary for students to be able to benefit from independent concordancing, a strategy which I argue can supplement traditional methods.

2.1 Foregrounding Vocabulary in English for Academic Purposes

The number of international students and other English learners in American
universities continues to rise as a university education is seen as a key to professional and
economic success, and as English maintains its status as the lingua franca throughout
much of academia internationally (Cobb and Horst 2001). The backgrounds, goals,
interests, and needs of these students are diverse, and their time spent in intensive English
programs preparing for English-medium university study is often limited. Since time is of
the essence, instructors must provide their students with knowledge of a core set of
academic vocabulary that is commonly found across disciplines, and with strategies that
will allow them to acquire and utilize the specialized and technical vocabulary that is
particular to the discourse community of their chosen field (Gardner 109). Moreover, it is
absolutely essential for instructors to emphasize the connections between the register of a
discipline or rhetorical situation and the lexical items that are most important in that
context. Often, "speaking the language" of a discourse community entails more than just
communicating in English; it requires the appropriate use of a specific lexis (Swales 26).

As growing numbers of language learners enrolled in English-medium
universities in the late twentieth century, their instructors became aware of the specific
needs these students brought with them (Santos 1988). Although second-language writers
shared many of the needs of students from English-speaking backgrounds, instructors and
researchers began investigating the error patterns and learning requirements that were
unique to this growing population of students. A key moment in this process was Terry

Santos's 1988 study which demonstrated how gaps in language learners' lexical

knowledge could be extremely detrimental to their professors' perceptions of their

writing skills.

In Santos's article "Professors' Reactions to the Academic Writing of Nonnative-

Speaking Students," she reported the findings of her study, in which 178 professors were

asked to rate two 400-word essays written by NNSs (69). Overall, the professors who

were surveyed found the student writing "highly comprehensible, reasonably unirritating,

but linguistically unacceptable" (76). Of all the problems present in the papers, lexical

errors—errors in word choice—were considered by the professors to be the most serious

(84). Santos infers from her data that although professors attempted to evaluate the

content and language of the students' writing independently, lexical errors are the site at

which this approach became impossible due to the way that these errors interfere with the

effective communication of ideas (84). She explains, "When the wrong word is used, the

meaning is very likely to be obscured" (84). From her findings, Santos concludes that

writing instructors of NNSs need to place "greater emphasis on vocabulary improvement

and lexical selection" (69). This idea is still relevant for teachers and students today.

Indeed, although Santos had drawn attention to the connection between lexical

knowledge and effective writing in 1988, an article published in 1999 referred to

vocabulary as a "still-neglected element essential for the second language writer" (Lowry

7). In the article "Lexical Issues in the University ESL Writing Class" Mary S. Lowry

argues for better teacher awareness of the issue, explaining that it had often been

overlooked in both language learning research and ESL writing instruction (8). She

observes that the role of vocabulary in writing is often neglected in books about teaching

ESL writing, and that the vocabulary-writing connection is similarly slighted in texts on

teaching and learning vocabulary (8-9).

Lowry argues that many EAP programs lack an explicit focus on vocabulary

awareness. She points out that teachers assume that writers at advanced levels will

already have most of the vocabulary they need in order to employ the proper terms in

academic writing situations, which is often not the case (8). Citing earlier research, she

points out common error patterns that occur based on limited lexical knowledge: "word

form errors" (e.g., the painting is *beautifully*), "inappropriate use of words from the oral

register" (e.g., "gonna"), "confusions between similar words" (e.g., there are many

people *sanding* on a beach), "preposition errors" (e.g., we arrived *to* our destination), and

"markedly poor control of abstract language" (e.g., *although* the results are inconclusive,

but we can still learn from this experiment) (8, examples mine).

Lowry's article establishes the exigence for more attention to vocabulary in

writing instruction. She explains how lexical errors are often addressed in the process-

oriented writing classroom: too little and too late (9). She explains that ELs need more

explicit focus on vocabulary, and that without this instruction they will encounter many

problems due to limited lexical knowledge. These problems can include excessive

paraphrasing (often approaching plagiarism), "choppy sentences with markedly poor

coherence," thesaurus-dependency, non-idiomatic constructions, and errors that reveal a

gap in a student's knowledge of a term's grammatical behavior (10). She points out that

these lexical errors demonstrate the complexity of what it means to "know" a word. She

cites vocabulary expert Paul Nation's (2001) criteria for word knowledge to elaborate upon this complexity. Nation's criteria consist of understanding: denotation, part of speech, frequency, register, collocations, grammatical behavior, connotations, associated terms, shades of meaning, derivations, spelling, and pronunciation (31-45). In light of the complexity of vocabulary learning and the problems that stem from limited vocabulary, Lowry argues that instructors "must be committed to foregrounding lexical issues as often as possible in instruction and via specific assignments" (13). Thus, Lowry argues that a focus on vocabulary is critical in L2 academic writing instruction.

Nearly contemporaneous with Lowry's 1999 article, Averil Coxhead utilized a corpus of academic writing to develop the Academic Word List (AWL), which consists of the headwords of "570 word families that account for approximately 10% of the total words (tokens, or individual instances of words) in academic texts" ("Word List" 213). A headword is a *lemma* or basic form of a word which serves as the main entry in a dictionary or word list. The other words in the same word family would signal an occurrence of the headword when calculating lexical frequency. To illustrate, the word "understand" would be considered a headword or lemma (the word used as a main entry in a dictionary), while inflected or derivational forms such as "understanding," "understood," "understandable," "understandably" would be counted as instances of the headword when calculating how often it appears in a text or group of texts. Coxhead's 570 word families represent the most common vocabulary items in academic writing. The importance of this list in teaching academic vocabulary to language learners—and thus

on the teaching of vocabulary for the purposes of university writing—cannot be overstated.

Coxhead's list was created in direct response to two factors. The first is the idea that lexical features vary across registers and that the academic registers used within specific disciplines share many terms from a common lexis. Throughout academic writing from many disciplines, terms such as "indicate," "demonstrate," "method," "occur," and "theory" are ubiquitous, and these terms are more likely to be found in academic writing than in other registers, such as fiction. The second factor contributing to Coxhead's development of the AWL was the availability of a sufficiently large corpus— the one Coxhead used consisted of approximately 3.5 million words—representative of the kinds of texts encountered in a variety of academic disciplines, coupled with the technology required to analyze this corpus (214-17). Coxhead's list has obvious utility for vocabulary instruction and theory, so most scholarly work on these topics after its publication have referenced or used it in some way. Many teaching tools aimed at preparing language learners for the demands of English-medium university coursework continue to utilize the Academic Word List. It is important to note that an updated list serving a similar function was developed in 2013 by Mark Davies and Dee Gardner: the Academic Vocabulary List (AVL), which is based on a far larger academic corpus of 120 million words.

Coxhead has continued to make important contributions to the field of ESL vocabulary instruction, and she has emphasized the connection between vocabulary knowledge and effective academic writing. In a 2007 article, "Preparing Writing

Teachers to Teach the Vocabulary of Academic Prose," written by Coxhead and Pat

Byrd, the authors summarize and synthesize research findings on the topic of analyzing

grammatical and lexical features of academic discourse and how teachers can use the

insights from this analysis to provide direct vocabulary instruction and opportunities for

student practice. Coxhead and Byrd explain the types of grammar and vocabulary

instruction that must be emphasized in an ESL writing classroom in order for learners to

match lexical and grammatical convention expectations of specific academic genres or

registers (130). They distinguish between English-for-general-purposes approaches to

vocabulary and approaches informed by study of language-in-use in particular contexts,

pointing out that lexical, grammatical, and lexicogrammatical features are characteristic

of particular registers associated with academic discourse communities, and arguing that

teachers can prepare students to communicate effectively within those communities by

drawing attention to those features and allowing students to practice using them in their

own writing (132). They employ David Corson's concept of the "lexical bar," or the

vocabulary hurdle that learners need to overcome in order to participate meaningfully and

successfully in an academic field, explaining that mere exposure and chance are

insufficient for the complex and challenging tasks of vocabulary acquisition (132). The

article also emphasizes the utility of Coxhead's Academic Word List and examines the

implications of that general-purpose list on writing pedagogy, pointing out that

vocabulary instruction must be an explicit and carefully planned aspect of writing

pedagogy (132-33).

Coxhead and Byrd also establish and explain the connections among academic reading, vocabulary, and writing (133). The authors suggest that through careful reading with attention to text content and, more importantly, grammatical and lexical features, students can learn which terms and structures are important for writing in an academic field (133). They point out that many EAP programs designate vocabulary instruction as the responsibility of reading teachers, but argue that the connections among reading, vocabulary, and writing need to be emphasized in the writing classroom as well (133). Thus, lexis can be seen as a bridge between receptive and productive academic language skills, a detail which I will argue can be exploited through student corpus consultation.

Coxhead and Byrd complicate the notion that lexical and grammatical features are distinct categories by pointing out that the line between the two becomes blurry when we consider the many complex—and crucial—multi-word items that students face (134). They classify these multiword lexicogrammatical items into types: phrasal verbs (e.g., "agree to"); idioms (e.g., "red herring"); common collocations, or words that tend to be found near the target word (e.g., "father" is frequently found near the word "mother," as in "mother and father," so "father" is a high-frequency collocate of "mother"); key terms that typically demonstrate common patterns of grammatical behavior (e.g., whether a verb can be followed by a *that*-clause, or whether a verb is transitive or intransitive); lexical bundles that transcend register boundaries (e.g., "in order to"); and semi-fixed verb sequences (e.g., "may well be") (135-38). Applied linguists and ESL scholars have continued to argue that attention to these multi-word items is fundamental to both vocabulary acquisition and academic writing (Lewis 2000, Gardner 2013). However, a

comprehensive analysis of multi-word items is beyond the scope of this current research, which instead focuses on collocational patterns of *single-word items*, or how individual words fit into larger chunks of language.

In addition to focusing on the vocabulary-writing connection from a linguistic and pedagogical perspective, Coxhead's work has also examined language learners' views on the issue. In her 2012 article "Academic Vocabulary, Writing and English for Academic Purposes: Perspectives from Second Language Learners," Coxhead reports findings from a small, qualitative study of 14 university-level English language learners who were interviewed about their lexical choices after they completed a short, integrated reading/writing task. Coxhead argues that student interviews and her analysis of their writing samples demonstrate that these advanced students are very aware of the importance of their lexical choices on whether their writing will be deemed acceptable within an academic discourse community (142). She finds that the students also demonstrate awareness of audience and context and considered these factors in the lexical choices they made while writing (142-43). Although the concept of "register" is not well-defined in the article, Coxhead does note that academic registers shape teachers' lexical expectations—and hence students' lexical choices (137-38).

The students who participated in the study reported that while completing the writing portion of the activity, they knew they needed to use technical, professional, and academic words and noted that receptive skills (e.g., those used in reading or in listening to lectures) were very important in their determination of the key terms associated with an academic field (139-43). Students were aware that the lexis of academic reading,

lectures, and writing was different from the lexis of casual speech, and they also noted

that they might pay more attention to certain new terms if they thought the terms would

be used in their area of academic or professional interest (139-42). The findings of this

study bolster Coxhead's argument that vocabulary and register need to be foregrounded

in EAP programs preparing students for success in various academic disciplines.

　　　　Echoing Coxhead's 2012 qualitative study of student perceptions of lexical

choices in university writing, a 2015 article by Nicole Brun-Mercer and Cheryl Boyd

Zimmerman presents findings of a small study that examines students' decision-making

processes as they select vocabulary items to use while writing. In the article, "Fostering

Academic Vocabulary Use in Writing," the authors explain their study. In the study, nine

advanced learners from multiple language backgrounds were asked to read an essay of

approximately 800 words; write 250 words in response to a prompt on the reading while

using the original text, a dictionary, and the internet as resources; complete a survey

about vocabulary strategies; and then—after experienced ESL instructors had assessed

the writing and identified appropriate and inappropriate choices regarding academic

vocabulary use—respond to interview questions about specific words they used in their

essays (Brun-Mercer and Boyd Zimmerman 133).

　　　　Based on the students' responses to their questions, the authors conclude that "all

nine participants, regardless of performance on this essay or previous standardized tests,

recognized the importance of academic vocabulary in composition" (134). They also

found that the students were aware of the differences between academic and non-

academic registers, and that this awareness helped them to decide which words to use in their writing (134).

Although the study showed that students understand the general concept of the connections between register and effective word choice in academic writing, its authors pointed to data suggesting that this understanding does not necessarily lead to effective application of those ideas in the students' academic writing (134-35). Some of the students in the study were unable to correctly classify the terms they used according to their register even though they understood that they were expected to use academic vocabulary in academic writing; in other words, some students chose to use non-academic words (e.g., "stuff") that they had misidentified as academic due to exposure to these terms in classroom settings (135). On the other hand, some students who had learned, practiced, and effectively utilized individual terms previous to the study had a high degree of confidence in using those terms in the writing task (135-36). From these observations, Brun-Mercer and Boyd Zimmerman conclude that due to the complexity of language learners' tasks in choosing appropriate lexical items for university writing assignments, teachers need to promote vocabulary acquisition and greater attention to register (141).

I have sketched out a brief and admittedly limited survey of some of the research on the impact of ELs' lexical choices and register awareness on the quality of their academic writing. Researchers have been contributing to this field of inquiry from a variety of valuable perspectives. Santos examines the topic from a real-world, contextualized position by soliciting the perspectives of university teachers who do not

necessarily specialize in teaching language learners. Others, like Lowry, analyze students' lexical errors and used the findings to encourage teachers to have a better understanding of the importance of the vocabulary-writing connection and adopt better-informed pedagogical approaches. Coxhead contributes to the field in many valuable ways, both in terms of analyzing academic language to determine the essential words that all university language learners need, and, with Byrd, in continuing to draw out the nuances of vocabulary in academic writing so that the importance of foregrounding vocabulary instruction in writing classrooms is clear. Coxhead, Brun-Mercer and Zimmerman demonstrate the value in interviewing and surveying language learner university students in order to gain an understanding of these issues from a learner's perspective. Together, these complementary approaches clearly demonstrate that the issue is as complex as it is crucial to student success in the university and in achieving their personal and professional goals. This M.A. Project aims to contribute to this pursuit by providing a practical approach to training students in using corpus data to enhance their acquisition of specialized vocabulary.

2.2 Principles for Teaching Vocabulary in English for Academic Purposes

The idea that ESL writing instructors must help students continue to expand their personal storehouses of academic vocabulary, understand the effect of register on lexis, and employ independent strategies for learning and appropriately using new words and multi-word items is well-established in the field of ESL pedagogy (Gardner 2013). What is less clear is how to determine the best approaches for translating these insights into

curriculum design, learning materials, lesson plans, and activities that will allow students to develop the knowledge and skills necessary for them to become effective independent vocabulary learners when they begin English-medium coursework. Fortunately, many language pedagogy experts have proposed guiding principles and classroom strategies that will help. See Appendix C for a convenient reference list which I have compiled from several key scholars' recommendations. These ideas are discussed below.

First, articulating vocabulary–related goals can be valuable, both for students and as an exercise in establishing a rationale and framework for curriculum design and lesson planning. Mary Lowry (1999) suggests setting the following six goals:

1. promoting student awareness of "the importance of intentional study for becoming a good writer"

2. providing independent vocabulary study strategies

3. tying vocabulary study to students' writing

4. "providing guided practice"

5. "familiarizing students with a selected body of academic vocabulary that will be useful in writing for various content area classes"

6. establishing a "response mechanism for instructor feedback…and answers to students' questions about words" (13-14).

In addition to these suggestions, Coxhead and Byrd recommend that students should be encouraged to "expand their academic vocabulary...[b]ecome aware of the differences between academic vocabulary and the words they use in conversation with friends...learn how to sort through words...[u]nderstand that 'learning a new word' means more than

memorizing a synonym or dictionary definition" and "[u]nderstand that 'learning a new academic word' means learning significant collocates or recurrent lexical sequences in which the new word is embedded" (143). Collocates and lexical sequences both refer to the words that can be seen surrounding the target item. For example, Coxhead and Byrd point out that the word "single" is a collocate that frequently precedes the word "mother" (136). The phrase "single mother" would thus be considered a two-word lexical sequence.

After determining students' academic vocabulary needs and establishing course vocabulary goals, instructors need to think about how to provide effective vocabulary instruction. A comprehensive explanation of ESL vocabulary instruction is, of course, beyond the scope of this M.A. Project, but it will be valuable to establish some guiding principles, such as *direct instruction*, *register awareness*, and *collocation training*.

In *Exploring Vocabulary: Language in Action,* Dee Gardner promotes a model of direct instruction that prepares instructors to enable their students to become independent vocabulary learners. Gardner argues that instructors should help students deal with lexical issues metacognitively and help them practice strategies for learning new vocabulary. Gardner's model includes "conceptualization" (i.e., mentally associating terms with their meanings); "form and meaning practice" (e.g., studying vocabulary flashcards); "context-based word-learning strategies" (i.e., using the sentence or paragraph containing the unfamiliar word to guess its part of speech and meaning); dictionary definition training; "morphological awareness raising" (e.g., learning to make inferences based on prefixes and suffixes); and "collocation training" (i.e., determining which words are often used in

conjunction with target items)" (118). Moreover, Gardner recommends explicitly focusing on register features in vocabulary instruction, and using an English-for-Specific-Purposes (ESP) approach to determine the terms and texts with which individual learners need to practice (81).

Gardner maintains that an ESP approach to vocabulary differs from a general-purpose approach in that it aims to help learners acquire the terms that appear more frequently within particular professional fields or academic disciplines than they do in general contexts (14). If teachers are aware of the registers in which their learners need to read and write, selecting texts and designing writing assignments can be done much more effectively, Gardner argues (64). These instructional approaches which promote students' independence in dealing with new vocabulary can be particularly effective for advanced learners who are preparing to face unfamiliar academic and discipline-specific vocabulary.

Nicole Brun–Mercer and Cheryl Boyd Zimmerman also provide helpful classroom applications in their article on academic vocabulary in writing. They encourage teachers to explicitly teach a vocabulary word's register (e.g., formal, slang, or technical) when teaching the word to the class, and they suggest that offering synonyms of the new word from other registers can help students understand the differences between academic and non-academic words (138). They recommend teaching students how to use learner dictionaries—specially designed dictionaries with simplified definitions that are easier for language learners to understand—that list words' registers and whether they are academic words or not (138). They suggest "text-

correction…activities in which learners find words in the wrong register and find replacements" (138). They also recommend providing students with a small repertoire of very common academic words and phrases, including logical connectors like "on the other hand," arguing that students will feel more confident in expressing their ideas in writing when they have these items to fall back on (138). Most importantly for our purpose here, Brun–Mercer and Boyd Zimmerman encourage teachers to find ways to provide multiple exposures to target words in contexts that reveal how the words function (138). I argue that concordancing can fast-track these exposures to authentic use of key words.

Ultimately, teachers must prepare their EL students to write in English at the university level by encouraging them to make the connections among register, reading, and vocabulary. In my own teaching practice with advanced learners in an IEP, I have found that one way to achieve this is to assign independent vocabulary journals based on authentic readings from within the students' major field of study. Advanced level students can be instructed to locate an introductory textbook from the discipline in which they plan to major. They should establish a weekly reading goal, and the instructor should require them to keep track of a certain number of new vocabulary words each week. The students should fill in the journal with many details about each new word (see Appendix B): part of speech; definition; collocations; example sentences; words grouped within the same word family as a target word (e.g. "stratified," "stratification," "stratify," "stratifying," "stratifications," and "stratifies"); mind maps linking associated terms; first-language (L1) translation; and even illustrations (adapted from Schmitt and

Schmitt's vocabulary flashcards). The instructor can monitor weekly progress and make

sure that students are locating the definitions that are appropriate for the terms as they are

used in the students' chosen disciplines. The instructor can also help students identify

words that are more common across multiple disciplines (i.e., AWL words) and allow the

students to share those words with the class. I have found that students leaving the IEP

program after doing this activity report feeling more confident in their ability to navigate

their future coursework.

A consensus has not yet been reached about which vocabulary and register

teaching strategies will be most effective for improving the quality of student writing.

Perhaps the best approach for an ESL writing teacher at this point is to experiment with

different types of lessons and activities dealing with these subjects and evaluate lexical

choices in student writing carefully in order to determine what is most effective. In my

experience in working with learners at advanced levels, these students tend to have strong

metacognitive awareness of vocabulary acquisition strategies, so discussing their choices

and approaches with them can be very revealing. Although the vocabulary demands that

will be placed on ELs as they begin writing in various disciplines in universities are

complex and challenging, teachers who make a conscious effort to foreground vocabulary

instruction in the EAP classroom will be able to provide the most useful lexical items and

promote the independent learning strategies that learners need, thus improving their

students' chances of success.

2.3 From Receptive to Productive Vocabulary Knowledge

In the university environment, *receptive* skills (i.e., listening and reading) are connected to *productive* skills (i.e., speaking and writing), and both of these domains depend heavily upon lexical knowledge. As Randi Reppen states, "In academic classes…learners are expected to read material and then to be able to write or speak about the material" (22). The degree to which a student has mastered the key terms found in that discipline's readings will largely predict the student's ability to produce those terms accurately and effectively in their own writing.

Research on moving students' knowledge of target words from the receptive domain to the productive is relatively limited (Lee and Muncie 2006). In Lee and Muncie's research study exploring this area of vocabulary acquisition, 48 intermediate ESL secondary students were provided several exposures to target items in context, then instructed to engage the terms through a variety of in-class activities, and ultimately assessed on their ability to accurately employ the target words in three composition tasks, which were given at different intervals and in which varying amounts of direct vocabulary support were provided. What Lee and Muncie found contradicts Stephen Krashen's Input Hypothesis (1989), which posits that learners' incidental encounters with new terms in context are sufficient for vocabulary acquisition. Their study demonstrated that more direct means of scaffolding, such as "teacher elicitation, explicit explanation, discussion and negotiation, and multimode exposure to target vocabulary (i.e., opportunities to interact with the terms in a variety of contexts) more effectively

increased learners' use of target vocabulary" than incidental exposure to new terms in readings (314).

Thus, this study's implications for language learners in English-medium courses (given their lack of access to comprehensive vocabulary support from instructors) is that additional exposure and intentional, repeated, engagement with target words, when combined with practice in their use, will make it more likely that students will be able to produce them in appropriate circumstances. Additionally, one of the study's findings that has potential to be exploited is that the participants' long-term ability to recall and produce *lexical phrases* was greater than their long-term retention and production of *single-word items* (311). This finding highlights the importance of learning how new vocabulary terms commonly collocate with other words, an area to which I will now turn.

2.4 Collocation and Multi-Word items

Some applied linguists argue that when target words are learned in lexical chunks (i.e., with the target word surrounded by common collocates), a great deal of valuable information about their use is simultaneously learned, such as grammatical behavior, common syntactic patterns, and even how the word functions rhetorically (Lewis 2000). For example, searching the Corpus of Contemporary American English (COCA), a collection of more than 560 million words from authentic texts across a range of registers, for collocates of the term "stratification" reveals that the word is most frequently collocated with the word "social," and that this collocation occurs roughly 5.5 times more frequently than the second most common collocate, "economic" (COCA).

Furthermore, the corpus reveals that these collocates are usually located to the left of the target word (i.e., *left collocates*), while the third most common collocate, "system," tends to fall to the right of the target word (a *right collocate*). A conventional approach to vocabulary acquisition has been to encourage learners to use context to infer the meaning of a new word; when learning lexical chunks and collocates, a learner can acquire a word *and* some of its context in a single, larger unit of meaning.

Some, such as second-language vocabulary scholar Michael Lewis, would argue that, when learning an additional language, it is more valuable to learn lexical chunks rather than individual words (Lewis 2000). Michael Lewis's "Lexical Approach" promotes the learning of vocabulary—and collocation in particular—as the core component of language acquisition (Thornbury). Lewis argued that much of "a learner's mental lexicon…consists of multi-word items rather than individual words. The higher the student's level, the more this is true" (150). However difficult it might be to test the validity of Lewis's claim, many applied linguists recommend collocation training as a valuable vocabulary learning strategy (e.g., Gardner 2013). EAP and vocabulary scholar Laura Gavioli (2001) states, "[r]esearch shows that when learners engage in language learning and produce language, they do not do so by combining words and morphemes according to complex grammatical rules; rather 'we produce most utterances using multi-word chunks which we have stored as wholes'" (cited in O'Sullivan and Chambers 52). This suggests that a vocabulary-learning strategy which focuses on learning lexical chunks might more closely resemble the way that languages are learned naturally, outside the classroom environment. Such an approach might facilitate greater retention, fluency,

automaticity, and accurate production of new lexical items. Moreover, George Woolard, a proponent of collocation training, claims that

> the learning of collocations is one aspect of language development which is
> ideally suited to independent language learning….Collocation is mostly a matter
> of noticing and recording, and trained students should be able to explore texts for
> themselves. Not only should they notice collocations in texts they meet, but more
> importantly, they should select those collocations which are crucial to their
> particular needs (35).

This leads us to the question of how to explore these rich and meaningful collocations.

Applied linguists have gained many of their insights about collocations and lexical phrases through corpus linguistics, an area that is having an impact on how students learn vocabulary. Because of the wide availability of easily accessible digital texts, linguists are now able to compile *corpora*, or large, systematic collections of samples of authentic language in use. The possibilities for research using corpora—and the potential benefits of that research—are too wide-ranging to list here, but one teaching application that has gained serious attention is corpus consultation as an independent vocabulary learning strategy (Gavioli 2005; Reppen 2010). I argue that this approach is an ideal fit for a NNS who needs deep knowledge of discipline-specific vocabulary words in order to produce them accurately in speaking and writing.

2.5 A Review of Literature on Student Use of Concordancing

To learn how vocabulary items are used in authentic contexts, a language learner can use a concordancer, which is "a computer program that is able to search rapidly through large quantities of text for a target item (morpheme, word, or phrase) and print out all the examples it finds in the contexts in which they appear" (Johns 1988, quoted in O'Sullivan and Chambers 50). Student use of concordancing began as early as 1969, and linguists began promoting this as a viable learning tool in the 1980s (O'Sullivan and Chambers 50). In the literature, support for student use of concordancing has fluctuated since then. Proponents have argued for its potential value while critics have claimed that not enough empirical studies demonstrating its effectiveness have been conducted (Boulton 2010). Critics also point out that the lack of student-friendly concordancing technology could lead to frustration for learners attempting to use the clunky tools currently available on their own (Boulton 2010).

A seminal article from the era when corpus linguistics seemed poised to change language teaching and learning dramatically is applied linguist John Flowerdew's "Concordancing in Language Learning" (1996). Flowerdew optimistically claims, "there are signs that concordancing has reached the stage where it is about to have a significant impact on the organization and practice of language teaching" (87). Flowerdew outlines the concordancing software and corpora available at the time of publication, points out the potential benefits of dedicated corpora for specific disciplines in ESP contexts, and

then lists possible applications for teachers and learners (88-93). It is his suggestions for learner applications that concern us here.

Flowerdew suggests that learners can use concordance searches for error analysis (to check their own phrases against those output by a concordancer), serendipity learning (students freely using the concordancer to explore any linguistic features that interest them), and inductive learning (94-95). In the kind of inductive learning facilitated by concordancing, he argues, a learner sees many of examples of a word used in context and then uses the pattern that emerges to build a theory or infer a rule about the target word (or some broader aspect of the language itself) (95-96). He credits applied linguist Tim Johns (1991) with coining the phrase "data-driven learning" (DDL) to describe this inductive, student-led approach to understanding lexical, syntactic, and grammatical issues, and DDL appears in much of the subsequent literature on the topic (96).

Flowerdew provides a section of caveats in which he anticipates and addresses possible objections. He explains that corpus size and specificity are worth considering in relation to students' goals and areas of interest, pointing out that the sources and subject matter of a given corpus will have a strong influence on the types of examples of words-in-context that it will produce; thus, a learner studying ESP should seek corpora tailored to those purposes (see section 3.1 of this M.A. Project) (98). He then presents possible criticisms which, in fact, do become perennial talking points among scholars who are less enthusiastic about independent concordancing for students (see Boulton 2010). These include the concern that "many of the concordance lines will contain language which is beyond the proficiency level of the learners," that "not all concordance lines may provide

enough context to make the meaning clear," and that students may become frustrated if they can't locate the kinds of examples of particular usages of words that they are looking for (Flowerdew 98). To address all of these potential problems, he explains that learner training is essential for productive student concordancing (Flowerdew 99). As a final caveat, Flowerdew notes that there is a need for more research involving actual students using concordancing, citing Ma (1993) as "probably the only detailed evaluation to date" at that time (99). Although both proponents and critics of the method would continue to bemoan the shortage of empirical studies on its use and effectiveness, the appeal of student concordancing has been strong enough to allow a modest but promising body of research literature to develop (Boulton 2010).

Some of this research centered on questions about how effective student use of concordancing and corpora actually is, or whether the performance of students using concordancing would actually improve significantly compared to students who were not. Cobb and Horst (2001) conducted a study on the effects of lower-intermediate and upper-intermediate learners' independent corpus consultation on the transferability of new lexical knowledge to both a definition task and to the comprehension of novel texts including the target words. Based on the differences between pre- and post-test scores, they found that although gains on the definition task were not significant, on the novel-text-comprehension task, low-intermediate students gained 13.47% (compared to 2.52% for the non-concordancing control group), and the upper-intermediate group gained 16% (compared to 5.66% for the control group) (328). Although this study focused only on comprehension of new words in authentic contexts (and not on productive skills), it can

be argued that improved comprehension is a necessary step toward accurate production of new lexical items in speaking and writing.

O'Sullivan and Chambers (2006), two linguists who work in the areas of corpus linguistics and writing pedagogy, conducted a study on the efficacy of concordance training. In the study, 14 undergraduate students from the University of Limerick were trained in using corpora over a three-week period. During this period, they participated in a total of three hours of lecture and six hours of computer laboratory research activities, and they were instructed to use independent corpus consultation strategies to correct their own errors in a writing assignment (here in French as a second language). The researchers found that the students' work improved, especially in the areas of word choice and preposition usage, and they posit that many improvements were due to the writers' ability to locate lexical phrases actually used by native speakers (NSs) and use them in place of their own less-idiomatic constructions, errors which often resulted from L1 interference or attempted word-for-word translations (56-61). Additionally, feedback from students who participated in the activity was generally favorable towards independent corpus consultation.

The attitudes and perceptions of the students actually interacting with concordancers were first studied qualitatively and quantitatively by Yoon and Hirvela, two scholars whose work focuses on teaching language learners (2004). Yoon and Hirvela noted the potential benefits of student corpus consultation, explaining that the method's "simultaneous focus on vocabulary, grammar, and discourse patterns provides second language learners the kind of target language input they especially need to

achieve high levels of proficiency as L2 writers" (259). They observed that within all the

literature to date at that point, the focus had primarily been on the technology itself, on

linguistic insights, on student performance, and on teachers' use of the tools (259). In

order to understand how students feel about corpus use in writing, to determine whether

students find the experience beneficial, and to explore the kinds of criticisms anticipated

by Flowerdew (1996), they conducted a study of 22 primarily East-Asian students from

15 different majors (mostly science-related) in an advanced ESL writing course at an

American university (263). Throughout the first four weeks of their course, the students'

composition teacher supplemented their writing instruction with several 20-25 minute

sessions on corpus consultation. The students were trained to use the Collins Birmingham

University International Language Database (COBUILD) Corpus to generate "prototype

strings," or syntactical patterns in which target terms are commonly embedded (e.g.,

article + adjective + *term* + verb), synthesizing collocation and collocate-frequency

information discovered in their analysis of concordance searches for target words and

later to solve "word-related problems they were having in their writing" (265). The

teacher began with demonstrations and direct instruction and gradually led the students to

consult the corpus independently (265). After 18 weeks in the course, the students

completed a survey about their feelings on corpus use and their opinions on the method's

strengths and weaknesses (266). Additionally, the researchers interviewed four of the

students (266).

The results of the survey showed that the students generally felt positive about

using corpus consultation as a tool for writing tasks (268). Overall, it was clear that

students felt more confident in their writing after learning how to use the concordancer to solve problems or research word usage (268). Students did not generally feel that the corpus helped them with reading or grammar, and the researchers explain that the students were probably not familiar with the idea of lexico-grammar (the complex overlap of, and interactivity between, vocabulary and grammar conventions) and thus the information gained through concordancing did not match their understanding of grammatical knowledge (268-69). Students reported that the level of difficulty in using the technology of concordancing fell in the middle range of neither too easy nor too difficult (269). Interestingly, the advanced students who participated in the study had less favorable opinions about the approach than the intermediate students did, in contrast to prior literature which insisted that only the most advanced students would likely benefit from corpus consultation; the authors attributed this to the intermediate group doing more hands-on and in-class corpus work than the advanced group (272).

Supplementing the data provided by their survey, Yoon and Hirvela's student interviews also elicited valuable findings. The interviewees reported that, after training, they had begun using corpus consultation as an L2 writing tool for tasks beyond the class in which they were trained, finding the tool especially valuable for checking preposition collocates and "learning common usage patterns of words" (276). Moreover, the students compared corpus consultation and dictionary look-up, and determined that each had their strengths but that "a corpus is more useful for learning how and where to put words in context" (277). Yoon and Hirvela conclude their article by offering caveats similar to Flowerdew's (1996), explaining that in order for the tool to be used successfully, proper

training is required, and that concordancing is not a strategy that every learner will embrace (278-79).

Following Yoon and Hirvela, many articles on student use of concordancers include some element of students' perceptions of their experience with the technology and their feelings about its value. This is reflected in language-learning and lexis researcher Alex Boulton's (2010) survey of 27 empirical studies of corpus use by students, which Boulton categorizes as Data-Driven Learning, whether the articles themselves use Johns' (1991) term or not (129-31). Although the range of approaches taken in the 27 articles examined was too broad, and the data they generated too heterogeneous, for an actual meta-analysis to be created, Boulton was nevertheless able to find valuable patterns in the information (140). He noted commonalities among the research questions that tend to be investigated on this topic, and he divided them into three categories: those exploring whether students can actually use the technology; those (like Yoon and Hirvela) examining students' and teachers' feelings about using corpora; and finally those that measure "outcomes of using corpora as a reference tool, usually for writing, error-correction, or translation" (130). Boulton explained that he was primarily interested in actual learning outcomes from the approach, but he notes that very few studies have anything directly to say on the topic (130-31).

Boulton found that overall, papers in this area are being published at an increasing rate (134). He also found that many of the studies are based on observations of a limited number of participants over relatively short periods of time, and that many of the positive results reported are qualified in some way, leaving, as he says, "a total of six studies with

unambiguously positive findings that meet the normal requirements of statistical

significance" (139-40). Nevertheless, Boulton echoed the cautiously optimistic tone of

much of the literature on DDL, pointing out that the research shows that learner response

to concordancing is overwhelmingly positive (140). For our purposes here, it is important

to note that Boulton found that the majority of the studies focused on lexical features of

language, generally involving larger multiword items, lexical chunks and collocations

(136). He argued that "this correlates with Johns' insight that DDL is most effective 'on

the "collocational border" between syntax and lexis'" (Boulton 136, quoting Johns 2002).

2.6 Section Summary

Although it is clear that more research needs to be done in this area and that

concordancing proponents' enthusiasm for the technology's potential utility for learners

must be approached critically, I still find that there are sufficient reasons to recommend

concordancing and corpus consultation to advanced learners moving from IEPs into

English-medium university coursework. First, concordancing allows students to access

vast amounts of authentic text which can reveal the ways that key terms are actually used

in context. Second, corpus searches can be narrowed to cover only particular genres or

disciplines of interest, eliminating search results that may be irrelevant to those genres or

disciplines. Third, after proper training and practice, concordancing is another valuable

reference or tool that can be used by learners independently when they are studying new

terms or investigating how to use them in their own writing. When used in conjunction

with a range of independent vocabulary learning strategies such as dictionary use,

translation, morphological awareness, and keeping a lexical journal, concordancing—with the rich collocational, idiomatic, lexicogrammatical and lexicosyntactic information it can provide about target words—can be a very effective and efficient means of acquiring the depth of knowledge of discipline-specific vocabulary necessary for learners to demonstrate competence in academic discourses.

CHAPTER 3: PRACTICAL APPLICATIONS: PREPARING ADVANCED
LANGUAGE LEARNERS TO INVESTIGATE VOCABULARY THROUGH
CONCORDANCING TECHNOLOGY

As we have seen, the acquisition of academic and discipline-specific vocabulary
is crucial to language learners' success in mainstream English-medium university
coursework, and it has been argued that corpora and concordancers can aid students in
the vocabulary-acquisition process. However, relatively few texts on training students to
interact with corpora exist. In this chapter, I address this gap in EAP pedagogy by
drawing from some of the resources on student concordancing which are currently
available. I begin by providing a set of criteria for educators to consider when trying to
determine which corpus (or corpora) and concordancing tools are appropriate for their
students' purposes. I then outline a teaching unit for training students to become
independent, corpus-aided language investigators. Although these ideas could be adopted
(or adapted) for use in diverse teaching contexts, for the purposes of this research I have
chosen to emphasize those that are most suitable for advanced ELs preparing to complete
their IEP coursework and begin their university studies in English.

3.1 Selecting Corpora and Concordancing Tools

For an ESL teacher, it is one thing to know that ELs can use concordancers to
search corpora for authentic examples of language in use; it is quite another to sift
through the overwhelming number of corpora and concordancers available to determine
which technological tools will best serve their students. The digital landscape of available
corpora and concordancers is constantly changing as new tools are developed for

increasingly specific purposes. Appendix D shows just a few of the free online corpora

that are available to teachers of English learners. Yet, for the most part, use of these tools

by language learners themselves is seldom foregrounded in their design. This results in an

unfortunate mismatch between a potentially valuable research technology and those

students and teachers whom it could benefit.

Teachers should consider the following questions when selecting a corpus to use with

their students:

- Large or small corpus?

- General or specific corpus?

- Written or spoken corpus?

- Corpus of native speaker or learner language?

- Pre-existing or instructor-compiled corpus?

- Monolingual or parallel corpus?

- Reference or monitor corpus?

Because corpora range in size widely, it may be necessary to first consider which

size of corpus would be best for student use. This task becomes complicated, however, as

notions of "large" and "small" in relation to corpus size change over time. The earliest

corpora that linguists compiled (Lancaster-Oslo/Bergen Corpus and the Brown Corpus of

Standard American English) were considered large at one million words in the 1970s, but

thirty years later, a one-million-word corpus was considered small (Gavioli, *Exploring*

Corpora 7). To put this into historical perspective, it is important to note that entering

and coding raw language data into early computers was a monumental task; in contrast, many of the texts included in today's corpora began their lives in digital formats requiring much less processing to use with a concordancer. As technology develops and the mountains of digital text available for analysis continue to grow, notions of corpus size become increasingly relative. Nevertheless, I argue that larger corpora will be more useful than smaller ones if teachers and students seek linguistic data that best represents the language use in a target genre or discipline.

Additionally, a corpus's level of generality or specificity is a key criterion for evaluating its utility (Gavioli, *Exploring Corpora* 7). As corpus linguistics scholar Costas Gabrielatos explains, "[i]n terms of content, corpora can be either general, that is, attempt to reflect a specific language or variety in all its contexts of use (e.g., the American National Corpus), or speciali[z]ed, that is, aim to focus on specific contexts and users (e.g., Michigan Corpus of Academic Spoken English)" (3). For the purpose of discipline-specific vocabulary acquisition, it is, of course, ideal to search corpora compiled from texts belonging to the students' target disciplines when possible, while results from a more general corpus can provide a valuable counterpoint to the results obtained from a specialized corpus. For example, Averil Coxhead created the Academic Word List by comparing academic texts to fiction texts and identifying the 570 most frequent word families that accounted for 10% of academic writing, but only 1.4% of fiction ("Word List" 213). Students researching the vocabulary of academic disciplines may be able to use a similar contrastive approach to compare the frequency of vocabulary between corpora to determine which terms are key within their disciplines.

In a similarly contrastive approach to corpus selection, in certain circumstances a teacher may opt to instruct students to compare data from academic native-speaker corpora with data from corpora consisting of learner language. Since the primary goal of corpus consultation is for students to find samples of language that demonstrate patterns of usage deemed acceptable by members of target academic discourse communities, native-speaker corpora should be a teacher's first choice. However, corpora consisting of written texts and transcribed speech from English learners around the world (e.g., the International Corpus of Learner English) can be useful as well. As these corpora contain error patterns (e.g., miscollocations, nonidiomatic usage, and preposition errors) typical of EL students, some linguists have recommended that instructors use NNS concordance lines with their students for various purposes, such as practice in error correction. As instructors help students use native-speaker corpora to locate examples of conventional collocations and usage, lines from learner corpora may serve as counterexamples.

Another important consideration when examining the varieties of language contained in corpora is whether spoken or written language is more appropriate to explore. Although corpora consisting of transcribed spoken English (e.g., the Michigan Corpus of Academic Spoken English) can offer valuable insights into authentic usage, a corpus of written academic English (e.g., the Michigan Corpus of Upper-level Student Papers) is more likely to contain contextual information that will enable learners to use lexical items in some of the most specialized modes of discourse (e.g., academic essays and articles in scholarly journals), which is the aim of this M.A. Project. Although there may be some discrepancies between the way that discipline-specific terms are used in

writing and in speech, I suggest that practice with these terms in reading and writing will transfer to listening and speaking more effectively than the inverse.

Another consideration in corpus selection is whether to use one of the many large corpora available or to compile a small, targeted corpus of selected texts. A number of digital concordancing tools allowing users to input digital text to create customized corpora are available. Some are only available for purchase (e.g., WordSmith tools and MonoConc), but Tom Cobb's *Text-Based Concordances* (available at www.lextutor.ca/conc/text) is free and allows users to create microcorpora by pasting up to 10,000 words into a text box or uploading up to 50,000 words in a file. Similarly, *WordAndPhrase* (www.wordandphrase.info), a free companion site to Mark Davies' COCA resources, offers a text analysis tool which sorts the words in user-entered text, providing a great deal of useful information on each word, such as definition, part of speech, synonyms, collocates, frequency by genre, and whether the terms are found in the AVL. These applications might be of special interest to teachers and learners, depending on their goals. For example, an instructor might demonstrate how an EL student might want to run all the readings from a particular course for an entire semester through this concordancer ahead of time in order to preview difficult vocabulary that may occur at relatively high frequency. Although it would be ideal to choose one pre-existing concordancer with access to a sufficient body of text in order to avoid overwhelming students with too much autonomy or too many tools, a brief introduction to self-compiled corpora and other vocabulary analysis tools like those available at the *Compleat Lexical Tutor* (www.lextutor.ca) and *WordAndPhrase* (www.wordandphrase.info) may help

illustrate key aspects of independent vocabulary acquisition, such as identifying specialized and technical vocabulary in a text.

A further consideration in corpus selection—and one which may not be obvious—is the language and dialect of the corpus. Corpora, not surprisingly, have been compiled from texts and speech in many different languages. Moreover, a teacher may even choose the variety of English they wish to investigate through corpus research (e.g., British or American). Clearly, a student seeking to learn to use English effectively in academic contexts should use a corpus of English texts connected to those contexts. However, a monolingual English corpus offers no support in a student's L1. To assist linguists, translators, and language learners, *parallel corpora* have been created.

Parallel corpora generally consist of parallel texts in translation, that is, texts in both their original language and translated into a second language. Examples of parallel corpora include the English-German Translation Corpus, the English-Norwegian Parallel Corpus, The Intersect Parallel Corpus (of English and French), and the large MultiUN, which is a multilingual corpus of 300 million words of United Nations documents translated into the six official languages of the UN (Arabic, Chinese, English, French, Russian, and Spanish). A parallel corpus allows researchers and students to observe acceptable usage, grammar, structure, and idiomatic features of the same text in two languages simultaneously. This is a very powerful approach to concordancing which can be useful in language learning and will probably continue to grow in importance in the future. I will discuss my personal experiences using a French-English parallel corpus in L2 academic writing in the following chapter. Despite the advantages presented by

parallel corpora, there are some limitations which make them less than ideal for the purposes outlined in this project. The most obvious one at the moment is that the number of texts that have been translated within any particular genre of writing will always be less than those that have not been translated. Another limitation for our purposes is that most IEP instructors are not teaching classes with one homogeneous L1.

Another consideration in corpus selection is whether a static or dynamic relation to linguistic material is preferable. Some corpora consist of fixed collections of language; these are called *reference corpora* (Gabrielatos 4). *Reference corpora*, effectively, are closed systems to which no new linguistic material is being added. *Monitor corpora*, on the other hand, are open-ended and, in theory at least, are being added to over time (4). The benefits of using reference corpora include clearly defined borders and a stable set of data regardless of the time of access. In contrast, the benefits of using monitor corpora include the possibility of including the most recent texts available and monitoring linguistic variation diachronically.

For the purposes of helping students acquire the vocabulary of their chosen fields as these terms are currently used, I argue that a current, dynamic, and frequently updated monitor corpus would present more benefits to more learners than a static reference corpus. In many disciplines, the content of the field, its linguistic conventions, and its lexis change over time—sometimes drastically or rapidly. Thus, the way language was used in a discipline decades ago can be markedly different from the way it is used today. Consider the semantic drift (or change in meaning over time) of a word like "liberalism" in philosophy and political science, and the rapidity of change in both content and

terminology in a technological field like computer science. In these and other disciplines, lexical changes can be evident in time periods as short as a few months. Therefore, in order to best prepare students for the lexical demands of their chosen academic and professional fields, I argue that monitor corpora are the ideal option.

Taking the above factors into consideration while evaluating the English corpora and concordancers currently available, I have chosen a large, monolingual, monitor corpus, the Corpus of Contemporary American English and its embedded concordancing tools, as the best option for introducing language learners to these technologies for several reasons. First, the COCA is currently a free resource. Registration is required, and occasionally searches are interrupted by splash screens encouraging users to upgrade to a paid account, but overall, students will have access to a great deal of lexical data for no cost. Second, by current standards, the COCA is very large (560 million words and counting) and contains a great variety of authentic language from both popular and academic genres of text. Third, in terms of levels of specificity, it is ideal because it allows researchers to focus on writing within academic disciplines, but it also contains a huge collection of general texts outside of those disciplines, thus providing the ability to easily contrast academic language use to more general use.

The COCA is organized by genre into broad "sections" (*Spoken*, *Fiction*, *Magazine*, *Newspaper*, and *Academic*), which are then further subdivided by year (or a range of five years) and category. The COCA divides its academic content—103 million words culled from around 100 academic journals—into nine categories: *education*; *history*; *geography and social science*; *law and political science*; *the humanities*;

philosophy and religion; *science and technology*; *medicine*; and *miscellaneous*. Not all of these subcategories allow for truly discipline-specific corpus research; for example, someone looking for examples of a term in use in sociology would find results from within all the social sciences *and* geography. However, advanced users can use the COCA's virtual corpus feature to generate a subcorpus of only those texts from within specific genres or date ranges which contain a word or phrase defined by the user.

In terms of overall design, features, user interface, and ease of use, the COCA is a good fit for advanced language learners preparing to enter English-medium universities. Considering the site's functionality and breadth of search options, its layout is relatively uncluttered and easy to navigate. COCA's integrated concordancer is accessible through a running top bar and allows users to research several aspects of a word's usage at various levels of specificity. Broadly, a user can move from left to right across the four categories of the top bar (*Search*, *Frequency*, *Context*, and *Context+)* to generate, respectively, lists of frequency data, collocations and their frequency data, key words in context (KWICs), and expanded context with detailed information about the source of the language sample in which the term is found. Granted, the vocabulary of the site itself is somewhat specialized (e.g., "frequency," "collocates," "strings," "KWIC," and "context,") and will require some unpacking as teachers help their students make the most of the COCA. There are also many search options available which may or may not be of interest to English teachers and learners, such as the ability to compare frequency and collocational patterns of a word and a similar or synonymous word, and the ability to track frequency over time. The sheer number of research options could potentially

overwhelm new users. With training and practice, however, students should be able to use the COCA to explore newly encountered vocabulary effectively.

3.2 Training Students in Concordancing

Since linguists and scholars first began suggesting that corpora and concordancers could be used directly by language learners, many have emphasized that these tools are well suited to *discovery learning* (Johns 1991). In this type of interaction with language, the learner takes an active role as a language researcher, carefully sifting through large amounts of text to make observations and draw inferences about the way words are used in authentic contexts; many argue that this type of interaction with language provides more meaningful learning of target vocabulary than traditional, textbook-based methods, and that this in turn will enable students to use the terms in question with greater accuracy in the future (Johns 1991; Bernardini 2004; Gavioli 2005; Boulton 2010) Moreover, discovery learning promotes learner autonomy and a democratized learning environment (Bernardini 27-28). On the other hand, much of the growing literature on classroom applications of concordancing also emphasizes that students must be guided carefully as they initially begin working with corpora (31). It can be concluded, then, that instructors should take a cautious and considered pedagogical approach toward concordance training in order for students to gain the most benefits from a potentially overwhelming technology.

To help move students from inexperience to autonomy in concordance-assisted lexical research, I propose employing a teaching method that moves from direct

instruction to guided practice and finally to self-directed learning. The approach to concordancing training I outline incorporates effective teaching practices such as activating students' prior knowledge, promoting a collaborative learning environment, allowing students to engage with the material in ways that connect with their personal interests and goals, and encouraging reflection and metacognition throughout the learning process.

<u>3.2.1 Goals and Objectives</u>

Teachers designing an effective curriculum for training students in concordance consultation must first establish the goals of the unit. If our goal is for advanced students to become familiar with corpora and concordancers as tools for fast-tracking vocabulary learning within their disciplines, then it is important to decide what this will look like in terms of observable objectives.

I suggest that upon completion of the unit, students should be able to accomplish the following objectives (see Table 1 below). They should be able to research a target word in a concordancing program and utilize the data located to explain several aspects of the word's meaning, grammar, and usage. They should be able to use context to determine what part of speech and which semantic sense (if more than one exists) of a word is being used in a given instance. They should be able to identify patterns of collocation (i.e., which words or structures are likely to precede or follow an instance of a word) and determine which patterns tend to occur most frequently. They should be able to identify miscollocations (collocations which deviate from conventional, idiomatic usage) in their own or peers' writing and use data from concordance lines to improve

these errors. A desirable but less essential outcome is for students to be able to research a target word by comparing its frequency of occurrence within a discipline-specific corpus to that of a less specialized corpus in order to answer the question "Is this word a key term within this field?"

Table 1: Proposed Outcomes of a Unit on Concordancing

Students Will Be Able To:

1. Locate the target item using the concordancer

2. Use context to identify part of speech

3. Use context to clarify meaning (or distinguish among possible
   meanings for polysemic terms)

4. Identify patterns of collocation

5. Determine which collocates occur most frequently

6. Identify any miscollocations in writing

7. Use concordance data to correct errors and non-idiomatic
   constructions in writing

8. Determine whether a term is a key lexical item within a given field

The goal of concordancing training is to provide students exposure to a language tool that can help them achieve their personal, academic, and career goals; the goal is (usually) not to train future scholars of corpus linguistics. Thus, we should bear in mind that the means—the exciting new tech tools—should not overshadow the ends—namely,

the acquisition of discipline-specific vocabulary for more accurate production within academic discourses.

3.2.2 Prerequisites

For students to benefit from exposure to corpus data and concordancing applications, it would be favorable if certain skills, knowledge, and dispositions were already in place. Ideally, advanced language learners preparing to begin English-medium university coursework will be equipped with the academic competencies that would make independent concordancing a viable tool. Specifically, before embarking on a unit on corpora and concordancing, a student should have a grasp of the parts of speech; an awareness of register and its effect on language choices; a basic of understanding of polysemy (the idea that a single word can have multiple meanings, sometimes only one of which is accurate for a given context); and, of course, some facility in using computers and the internet. The ability to work independently, a curiosity about language, and a tolerance of ambiguity are also factors which will promote a favorable learning outcome.

3.2.3 Activating Prior Knowledge

Concordancers are relatively esoteric digital tools, but some common digital tools can provide students a frame of reference as teachers introduce these tools and their functions. Almost all word processing software and web browsers feature a "find" function for scanning large bodies of text to locate target items. For example, Appendix E shows Microsoft Word's "find" feature locating instances of the word "need" in a draft of this project. Other tools, such as internet search engines like Google, analyze text that users input into a search bar, making predictions about related terms or frequently

collocated words based on patterns in the data available on the internet (Google's

suggestions based on a search for the phrase "I need" are illustrated in Appendix F).

These everyday digital tools operate on principles not unlike those used by

concordancing programs. A quick in-class demonstration of these applications will help

students understand the logic of a concordancer. Similarly, a word-cloud program can be

used to visually illustrate the concept of word frequency (see Appendix G).

Teachers should also tap into students' linguistic knowledge and emphasize the

importance of collocation in language learning. It may be helpful for teachers to point out

(or remind) students that much of their language learning consisted of chunks of more-or-

less fixed phrases or structures. Some examples of classes of multiword items are phrasal

verbs (e.g., "break up"), idioms (e.g., "beat around the bush"), discourse markers (e.g.,

"on the other hand"), stock phrases (e.g., "see you later"), and prefabricated strings (e.g.,

"the point is…") (Gardner 22). Advanced learners who are familiar with these kinds of

items will be primed to understand how these patterns of collocation create meanings that

go beyond the level of single words and facilitate communication when used in ways

deemed acceptable by interlocutors in a given context. Teachers can build upon this

conceptual foundation and help students begin to understand that low-frequency,

discipline-specific vocabulary items also tend to collocate with some words more

frequently than others.

3.2.4 Initial Exposure to Corpora and Concordancers

In students' first classroom exposure to corpora and concordancing, a teacher can

provide some background on the history of corpora, provide some examples, and

introduce the class to concordancing tools through lecture and demonstration. However, if the teacher instead begins by posing an example of a linguistic problem that the students are likely to face and then showing how technological tools can address the problem, the students are likely to be much more invested in the lesson.

One way for the teacher to illustrate the potential problem—a lack of familiarity with a seemingly important new vocabulary item—is to use a reading passage and locate a word that is unfamiliar to most of the class, low-frequency, and essential to the understanding of the text. For this exercise, the teacher should be careful to choose a text that does not offer enough context to aid in developing word knowledge (e.g., no parenthetical explanations or footnotes explaining the term). To illustrate, I have chosen the term "stratified." The following sentences come from the expanded context of the first result of a search for the term "stratified" in the COCA: "He said that 'the earth needs to be protected; humanity needs to be dignified.' The solutions include 'course correction' from industrial excess, **stratified** development and fossil fuel consumption" (COCA).

The teacher and the class can then work through a series of strategies to develop a better understanding of the word's meaning, allowing the class to review methods of researching newly encountered vocabulary while revealing the gaps that these methods leave open (e.g., about usage and collocation). It can then be suggested that these gaps can be filled through concordancing. Strategies can include using context to guess both part of speech and meaning, looking at roots and affixes for similarity to known words, and looking the word up in a learner dictionary or a bilingual dictionary. The class can

keep track of the new information that this process yields, then reread the passage containing the term and check guesses about the word's meaning. Ultimately, the process might reveal that although the term has now been researched to give a basic understanding of how it functions in one particular instance, this understanding offers little in the way of being able to transfer this knowledge to productive use in speaking and writing.

In order to learn more about how the term is used authentically, it is time to introduce the concordancer to the students. First, a basic context search can be conducted. The result will be a concordance line output that might appear almost random and probably overwhelming to the students upon first view, as seen below in Figure 2.

Figure 2: Screenshot of the KWIC "Stratified" in COCA (Davies)

In the example provided, the context results of a query for the term "stratified" yield a lot of information about the word. A quick glance down the third column reveals that most of these examples of the word are from academic (ACAD) texts. Some of the term's grammatical behavior can be inferred as well. For instance, as the class tries to make sense of the "decontextualized context" by locating common syntactical patterns, it should become clear that the word is being used here as either an adjective or a verb. Students familiar with passive constructions may even notice that the word is often used in this way as well.

Perhaps most importantly, the term's patterns of collocation will become evident. In the 20 concordance lines in Figure 2, lines two and three both show the collocation "highly stratified." The words that fall to the right of the KWIC (i.e., "right collocates") provide even richer information. In just 20 lines, "stratified by" (+ noun phrase) occurs three times, "stratified analysis" occurs three times, and "stratified according to" (+ noun phrase) occurs seven times. Astute student-researchers may even notice that the noun phrases following both "stratified by" and "stratified according to" lack articles and other determiners.

Whether these insights are pointed out and explained directly by the teacher or noticed and shared by the students, it should be clear that a great deal of information about how the term is used in academic texts is now readily available—and that no other word-research strategy would be able to provide this kind of information.

As students begin to notice a few patterns, the teacher can demonstrate the collocate-frequency search feature of the COCA concordancer. This powerful tool allows

students to view the common collocates of a target word, ranked in descending order of frequency. For fine-tuning searches, users may set the range which the concordancer will search, from zero to four words to the left as well as the right of the target item. In Figure 3 below, the top 20 results for both left and right collocates of the term "stratified" are ranked by frequency.

Figure 3: Screenshot of Collocate Frequency for the Term "Stratified" in COCA (Davies)

This search feature yields results that may seem overwhelming, but students can be encouraged to focus on the listed collocates, the frequency ranking in column one, and the bar graph column to the far right of the page. This will allow them to quickly ascertain the relative frequency of a collocate and understand how a word is most likely to be used in authentic contexts. In Figure 3, it is clear that "by" is the most common collocate of "stratified," and that, despite the prevalence of "according" in the initial

context search in Figure 2, that this collocate actually ranks eighth. Granted, this list

approach may be too analytical and decontextualized for some students, but by simply

clicking on one of the collocates, the user will be sent to another page of KWIC

concordance lines like those in Figure 2, but this time displaying only examples of the

key-word-plus-target-collocate. For example, clicking "by" generates a long list of

concordance lines containing the phrase "stratified by," allowing students to target the

chunk as a whole and observe its patterns of usage and grammar.

Another potentially useful feature of the COCA, and one worth demonstrating, is

the "frequency by section" function. This feature allows users to search an item (i.e., a

single word, part of a word, or multiword chunk) and see its relative frequency across

COCA's "sections," the five broad registers of English (*Spoken*, *Fiction*, *Magazine*,

Newspaper, and *Academic*) contained in the corpus. In Figure 4 below, the bar graph

makes it clear that "stratified" is used much more often in academic texts than in spoken,

fiction, magazine, or newspaper content. A teacher might point out that, given this data, it

is probably reasonable to surmise that "stratified" is a term worth learning, at least within

some academic disciplines. Additionally, the time-period data (e.g., 1990-1994), located

below the five sections, can show if usage of the word has become more or less frequent

over time. In the example, the target word is being used more frequently over time,

suggesting that it is a word with some currency and thus, again, worth learning.

Figure 4: Screenshot of Frequency by Section and Time Period for the Term "Stratified" in COCA (Davies)

A final useful aspect of the frequency-by-section feature is the ability to click on one of the five sections (e.g., *Academic*) and see how the target item is used within it. This allows the user to see which patterns of usage are more or less common depending upon the broad context in which the term is used. Perhaps more importantly for discipline-focused investigations of vocabulary, though, is that by clicking on the *Academic* section link in the first column, a frequency-by-subsection page opens, allowing the user to see an item's relative frequency across the nine academic areas delineated within the COCA (*History*; *Education*; *Geography and Social Sciences*; *Law and Political Science*; *the Humanities*; *Philosophy/Religion*; *Science and Technology*; *Medicine*; and *Miscellaneous)* and to see the term used in context within any of those areas only (see Figure 5 below). This is as close as the COCA gets to allowing truly

discipline-specific lexical research (if we set aside its "virtual corpus" feature, which is beyond the scope of this project). It is important to note that sections, subsections, and year ranges can also be defined on the main search page, allowing access to some of the features (such as collocates) discussed above, but that this feature is slightly less stable in my experience in using the COCA (i.e., sometimes it has worked correctly on my computer, while other times it has not).

SECTION	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT
ACAD:History	41	12.2	3.35	
ACAD:Education	125	9.4	13.24	
ACAD:Geog/SocSci	211	16.2	13.04	
ACAD:Law/PolSci	6	8.6	0.70	
ACAD:Humanities	31	11.9	2.60	
ACAD:Phil/Rel	21	6.7	3.12	
ACAD:Sci/Tech	63	14.1	4.48	
ACAD:Medicine	161	6.7	24.03	
ACAD:Misc	4	4.3	0.94	

Figure 5: Screenshot of Frequency across Nine Academic Areas for the Term "Stratified" in COCA (Davies)

Although by no means exhaustive, this introduction to the features of the COCA's concordancer should provide enough research strategies to encourage students to begin learning to use corpus data to enhance their vocabulary learning process.

3.2.5 The "Soft Version:" Students Working with Teacher-Generated Materials and Samples

After students have been introduced to concordancers—what they are, what they do, how they work, and how they can help students understand how words are actually used—the next step is for them to practice working with samples of concordance lines.

This relatively predetermined route through an investigation of authentic language data is what Costas Gabrielatos calls the "soft version" of classroom concordancing, in contrast with the "hard version" in which students use concordancers directly (10, 17). This section of the training could be built around printouts of KWICs, with a series of activities and questions for the students to complete.

A worthwhile point of entry for this section would be to promote students' inferential learning about collocation and patterns of usage. Teachers can provide lists of actual concordance lines for academic terms and ask students to note which words appear to the left and right of the term. This will help them form theories about common collocates. Very competent students can be prompted to search for other syntactic patterns—or even patterns in content. Such students might make more abstract observations, such as "This word is usually followed by a noun phrase," or "I'm seeing a lot of content that has to do with rocks," or "Every time I see the word used with these collocates, it seems to concern research methodology."

After repeating the process with several vocabulary items, a formative assessment from this section might use fill-in-the-blank or matching questions that test the students' ability to identify conventional collocations and avoid miscollocations. Ultimately, the students can be prompted to produce the terms and their collocates appropriately in new sentences.

3.2.6 Intermediate Stages: Finding Collocational Patterns with a Concordancer

The intermediate stages of training fall between the time students interact with uniform, teacher-prepared materials and the time when they begin using the technologies

to investigate usage patterns of new vocabulary on their own. During this phase, teachers can utilize a problem-posing pedagogical approach and guide students through a series of activities that can reveal the kinds of usage patterns that students will eventually learn to exploit on their own. Hands-on learning using computers is key at this point in the training.

The same activities outlined above in the section on teacher-provided concordance lines can be repeated during this section, with the added dimension of students learning to navigate the COCA's digital tools independently. Students can be provided lists of terms which they will be coached in researching. They can use context displays to make inferences about patterns of collocation. They can locate frequency data of various kinds and use this information to make predictions about the register and context in which they are likely to find the word.

To help students begin differentiating among the various words in a term's word family, wild-card searches can be introduced. In a wild-card search, a symbol (usually an asterisk) is strategically inserted into a search term in a position that could be occupied by several variant letter combinations. This is particularly useful in generating a list of words that share a common root but feature different affixes. For example, a COCA search for "stratif*" yields the following terms in descending order of frequency: "stratified," "stratification," "stratify," "stratifying," "stratifications," "stratifies," etc. Each one of these terms (or at least the most frequent) can in turn be viewed as a KWIC, and the forms can be contrasted with one another in terms of collocation.

Practical applications of concordance research can be emphasized during this section of the unit as teachers present students with a list of tasks that mirror the steps a student might go through when encountering new vocabulary in their actual studies. For example, a short, paragraph-long reading containing the new item can be a starting point for inquiry, and students can be directed to search for meanings in a learner dictionary. They can be asked which sense of a word seems appropriate for the context. They can be asked to identify conventional usage patterns and differentiate them from nonidiomatic constructions in fill-in-the-blank or multiple choice questions. The students can then be instructed to use the concordancer to locate collocational and usage information, and then apply this knowledge in a second round of formative assessment. Ultimately, students should be assessed on whether they can accurately apply the information discovered through this research when creating new sentences containing the target item embedded within conventional patterns of collocation.

3.2.7 Independent Student Concordancing

The culminating segment of a unit on concordancing should allow students to use the technology to perform their own vocabulary research. This inquiry should be student-led; that is, the teacher can ask the students to find real texts from within their future disciplines and read as much of the text as they can until they encounter a difficult new word they would like to investigate. Teachers can provide a template worksheet which guides students through the process of word-attack and corpus consultation, this time with no predetermined correct answers present at the outset. The students can compile a list of collocational, frequency, word-family, register, and part-of-speech data on the new

term, and demonstrate that they can write novel sentences utilizing the target term in at

least three of its most common collocational patterns. This whole process can be repeated

a few times so that the requisite technical and linguistic skills are reinforced.

3.2.8 Reflection, Sharing, and Next Steps

After students complete their independent research projects, they should present

their work and findings to one another, reflect critically and metacognitively on their

experience with concordancing technology, and attempt to determine whether they will

use the tools in their future academic work, and if so, how they will use them. This can

take the form of small-group discussions and brief individual presentations to the class. In

the small-group sessions, students can discuss the benefits and drawbacks of their new

vocabulary research tool, share strategies and experiences, and answer a series of guiding

questions that will reinforce the relevance of independent concordancing to their future

academic pursuits in English. Finally, students can each present their findings from the

research process on one word that they find especially interesting or useful. Action-

research-oriented teachers can compile the results of this process and use them later when

working with future students interested in the same disciplines.

3.3 Chapter Conclusion

In this practical section of the M.A. Project, I have provided an outline of criteria

for determining which corpus and concordancer will best serve students, and I have

mapped out an instructional unit designed to both demonstrate the value of independent

corpus consultation and allow students to begin utilizing this approach to assist them in

their own academic reading and writing. Ideally, this unit shows how concordancing

strategies can provide a much-needed bridge between the receptive and productive

domains of language. In the following chapter, I will share my initial experiences in

introducing concordancers to my advanced students in an IEP, as well as my tentative

attempts to use corpus research strategies in French, as I grappled with unfamiliar

vocabulary in my own second-language academic writing.

CHAPTER 4: EXPERIENCES WITH CONCORDANCING AS A LANGUAGE
TEACHER AND LANGUAGE LEARNER

4.1 Action Research: Introducing Concordance Data to Students in an Advanced Reading
and Discussion Course

I have been teaching English learners of all skill levels from very diverse
backgrounds in a broad range of contexts for nearly ten years. Like many teachers, I have
been reflecting critically upon my teaching practice and my repertoire of approaches and
activities in order to improve as a teacher over time. A recent development in my growth
has been a recognition of the centrality of vocabulary acquisition in the language learning
process, and I have been seeking ways to translate insights from research and theory on
this aspect of language learning into curriculum that will best serve my students' needs.
The following example may help illustrate my initial steps in integrating collocation and
concordancing into the vocabulary component of a reading course.

During my teaching in an IEP at a university in California, I instructed advanced
EL students in an academic reading and discussion course aimed at preparing
international students to study in American universities. During one eight-week session
of the class, I introduced corpus data and concordancing to the students for the first time.
The class consisted of two Japanese students, four Chinese students, and two Brazilian
students. There were four males and four females, and they ranged in age from their early
to late twenties. Student ability levels ranged from the low end—consisting of students
who had not yet tested into the advanced level but had opted to challenge themselves in

advanced courses—to the high end, consisting of students with near-native-speaker proficiency.

I designed the curriculum of the course to have a strong focus on academic and discipline-specific vocabulary, and I made it clear to my students that I wanted the class to equip them with the independent vocabulary research strategies that would help them succeed in their imminent university studies. We spent several weeks at the beginning of the course learning about the AWL (our program had not yet obtained AVL-based materials) and we proceeded to investigate the vocabulary of each student's prospective major throughout the course. This focus took the form of a lexical journal assignment discussed in Chapter Two and illustrated in Appendix B. I taught this class after becoming interested in collocation and concordancers but prior to engaging in the research which has informed this M.A. Project. This means that the approach outlined above in Chapter Three had not yet been formulated. Nevertheless, my experience teaching this class demonstrates some of the benefits and challenges of introducing concordancing and concordance data to advanced students.

While teaching this class and learning about students' experiences with an approach to vocabulary that was new to all of us, I employed an *action-research* methodology as defined by William Grabe and Fredericka Stoller (2013), two applied linguists who have published extensively on L2 reading. According to Grabe and Stoller, *action research* is a "type of teacher reflection, or teacher-initiated enquiry, during which teachers look critically at their own classrooms to improve their teaching and enhance the quality of learning that takes place there" (164). Thus, in this model, the teacher uses a

classroom as a site for gathering quantitative and qualitative data about practical issues in teaching and learning in order to explore a set of questions that the teacher has formulated about a particular topic (167). The process is systematic yet fluid and dialectical, meaning that research questions and even approaches to data collection may be refined in real time as the observed data shapes the teacher-researcher's theory and focus. The ultimate aim of action research is to generate knowledge of teaching and learning that will allow teachers to adopt practical solutions to problems which they and their future students are likely to face in similar contexts.

While teaching the IEP reading class, I initially wanted to research concordancing, and I wanted to know whether students would embrace concordancing as a vocabulary research tool and what the advantages and disadvantages of working with this tool might be. Based on my nascent understanding of research on student concordancing, I anticipated that concordance lines could provide collocational data that would enable students to better grasp how newly encountered terms are used. I also anticipated that the format of KWIC concordance lines might prove to be challenging to some students, possibly making inference learning difficult for some and thus making concordancing less valuable than I had hoped.

I introduced collocations, corpora, and concordancers to the class while outlining my expectations for their lexical journals, an extensive reading and vocabulary project on which the students worked for the full eight-week session. In this project, students chose an academic text from within their field of interest and explored it for the purpose of acquiring new vocabulary throughout the semester. Since two of the students were

already taking university courses within their fields along with their IEP courses, these students were able to choose from their actual course materials. All of the students were required to document ten new words each week in lexical journal entries. Each entry would contain the term, its part (or parts) of speech, a short definition in English, a translation into the student's L1, a list of other words from the same word family, three to five common collocations, and an example sentence. Additionally, in each entry, students had to choose any three of the following aspects of word knowledge to include in each entry: a mind-map graphic organizer showing associated terms, a visual illustration, phonetic pronunciation, register, frequency, synonyms, or antonyms.

To help students meet the requirement of finding examples of common collocations for their terms, I demonstrated the basics of the COCA. I also allowed students to quote the COCA when finding example sentences, although they were free to create sentences of their own if they chose. I pointed out that the COCA was also a good source for register and frequency information if they chose to include it in their entries. I demonstrated the COCA using a computer and projector in class during one class meeting, and reviewed the process in the same way twice in the first few weeks of the course. There was no guided, hands-on laboratory concordancing component during the session. This brief introduction to the COCA placed it alongside other useful online vocabulary resources like the *Oxford Learner's Dictionary*, *WordReference.com,* and *Thinkmap Visual Thesaurus.* I was able to assess the students' journal entries on a weekly basis, and this allowed me to qualitatively monitor the results of their engagement with these online tools.

During the class, another important aspect of framing the vocabulary component and introducing concordance data was one of our textbooks, Mikulecky and Jeffries' *Advanced Reading Power* (2007), which I was using for the first time. I had chosen this textbook for its emphasis on the importance of vocabulary in learning to read at the advanced level in English. The text features a 46-page section on strategies for building vocabulary while reading, and it also, conveniently, includes appendices listing the 2000 high-frequency, high-utility words of the *General Service List* as well as the 570 headwords of the AWL, which provided a useful point of entry for discussing frequency and register. I also chose the book because its readings come from authentic texts such as the *New York Times* or current introductory textbooks from various disciplines; this suited my goal of helping students to prepare to engage complex texts after they matriculated into university coursework. Most importantly for the purposes of my research on student concordancing, the text's vocabulary section concludes with a unit on collocations, and the unit guides students through the process of utilizing authentic concordance lines to make inferences about word usage.

The collocations unit in *Advanced Reading Power* introduces the concept of collocation, and then encourages students to find patterns in the way words collocate based on concordance lines. It ultimately asks them to apply this knowledge by matching words from a word bank to their appropriate collocates or filling in a missing word in an example sentence with a collocate that seems appropriate based on the patterns found in the concordance lines. Because the vocabulary work in this textbook was just one of

several concurrent segments of the reading course curriculum, the unit on collocation was completed over a period of two weeks, or four class sessions.

Using this two-pronged approach to student concordancing, i.e., guided practice with textbook exercises and independent searches of the COCA to meet the requirements of the lexical journal project, I was able to make observations about how students interact with concordance technology and data. These observations allowed me to reflect on my initial research questions and anticipated outcomes, which in turn helped me make decisions about how to integrate concordancing into future courses.

The textbook's approach to guided interaction with concordance lines in the text is as follows: students are provided twelve actual concordance lines using an AWL word as a KWIC; the students are then asked to list words or phrases found to the left of the target word (i.e., "left collocates"), and then note two patterns they can see from that information; the process is then repeated for words or phrases located to the right of the target word (i.e., "right collocates"); then, students compare answers with their partners and then the class discusses the words, phrases, and patterns that were discovered.

This part of the process led to results which generally confirmed my anticipated outcomes of students analyzing raw concordance data. Most students were able to identify patterns of collocates, usually both to the left and right, of target words when these patterns were evident. However, because the text employs real concordance lines that have not been edited or sorted to facilitate pattern recognition, sometimes patterns were harder to identify. For example, in the book's exercise using concordance lines for the word "perception," just one content word, "sense," is repeated in the twelve lines, and

the preposition "of" appears just three times to the right of the target word (Mikulecky and Jeffries 68). Fortunately, we were all working in class together, so students were relieved when the whole class and I agreed that a particular set of concordance lines didn't yield many particularly useful insights to one side of the target word. This helped me identify an important principle of independent student concordancing: sometimes the patterns just aren't there to be found (at least not readily), and a student might benefit from some tolerance of ambiguity in order to move forward while still seeing the value of the tools.

When it came to the more metalinguistic questions in the text, e.g., noticing that a phrasal verb tends to precede a specific target word, the results were more mixed and tended to vary according to students' ability levels and competence with grammar; this is to be expected. For those students who were able to successfully identify parts of speech, this aspect of the activities provided additional practice with grammar; for those who found this aspect challenging, this step seemed very frustrating. Because I was not also working with these students in their writing course, I was unable to track whether either type of student was able to apply these patterns in novel situations.

In terms of application of insights garnered from working with concordance lines to new contexts, results were again mixed. A typical exercise in the textbook asks students to choose from a 25-word bank of adjectives and match five adjectives that would appropriately precede four nouns—without consulting the concordance lines on earlier pages until finished. One problem we found with this type of exercise was that many of the adjective collocates seemed to apply equally well to several of the nouns.

Another problem was that, even after practice with concordance lines, students created collocations that were grammatically correct, and even made sense, but which for one reason or another would never be used by a native speaker. Without actual digital concordancing technology at hand to test these collocations, it was hard for students to see why these miscollocations were less appropriate than more frequent patterns.

Overall, my first experience using concordance lines with students helped me maintain a realistic perspective on the resource. It showed that, even when pre-selected and carefully organized in a textbook, concordance data can appear cluttered. It showed that there are multiple points in the concordance research process that can lead to frustration and possibly lead to students losing interest. Finally, it showed that there should be more guidance between the steps of pattern-noticing and applying these insights to novel situations. I concluded my notes on this experience by reflecting that although concordancing wasn't exactly the "magic bullet" I had hoped it would be for moving difficult vocabulary from the receptive to the productive knowledge domains, it did generate some useful insights for some students, making it one among many valuable tools that students can access when researching new terms.

This insight seemed to be corroborated by students' engagement with the COCA as they completed their lexical journals. On the whole, students reported being able to navigate the website on their own successfully, and they were also mostly tolerant of the ragged appearance of concordance lines, not minding that often it was difficult to locate beginnings and endings of complete sentences in them. Half of the students continued to use the COCA for the collocation and sample sentence requirements of the lexical journal

project throughout the semester. The other students opted to use collocation data from the *Oxford Learner's Dictionaries* and create their own example sentences rather than quoting them from concordance lines. To these students, it was simply faster and easier to use a single resource—a dictionary—for several aspects of word knowledge rather than interact with a concordancer.

One of the eight students really bonded with the COCA and became both adept in using it and enthusiastic about applying it in her future studies. She was one of the highest-achieving students in the class, and she was already taking a university course in her discipline. She had, in fact, chosen a particularly dense academic article written by her professor as her source for journal entry terms. Her entries showed that she was putting a great deal of effort into discerning collocational patterns, and the examples she included all seemed to demonstrate that she had learned several ways that the target words would be used in context by native speakers. This student's performance and reports about her continued use of the tool helped reinforce my wish to help make concordancing technology accessible to advanced students.

4.2 From the Language Learner's Perspective: Using Search Tools and Parallel Corpus Data in Extensive L2 Academic Writing

While completing my initial training to teach English to speakers of other languages, I had often noted that actually learning a foreign language in a classroom context helped concretize the theories I was encountering. For this reason, I recently put myself in the language learner's shoes once more. I chose to study French because I had studied it for two years in high school and felt that I was at a level comparable to many of

my own students' levels in English as they began studying in the IEP in which I worked. I wanted to identify with them and try to experience what they felt as they entered my classroom. In fact, to make sure I experienced the frustration and feeling of being overwhelmed by language that was slightly beyond my skill level—and the pride of success when I could sense my own progress—I opted to take a low-intermediate-level course. In this course, I found myself taking the initiative to use concordancing (and concordance-like tools) to solve problems I faced in utilizing new academic vocabulary appropriately.

The French course, like many foreign language courses, was a blended-skill class. In the 200 minutes we met each week, the instructor provided practice in reading, writing, listening, and speaking. There was decidedly little L1 support—which suited my goals perfectly. The class had a strong emphasis on grammar and discussion. Vocabulary instruction was more incidental, and the class usually learned new words that related to grammatical structures we were learning, or terms essential to comprehending readings or the theme of a unit in the course. My own vocabulary knowledge was somewhat limited, so I often struggled to express myself as I searched for words or constructed awkward workarounds when I didn't know them. Needless to say, I was not fluent in the French equivalent of the AWL words—except when a fortuitous cognate happened to exist in English.

In this context, I experienced L2 academic reading and writing firsthand. One assignment that the class was given was an extensive research writing project. We were required to write for ten weeks on any single topic related to French, and we were to

write roughly 300 words per week. Each week's writing was to contribute to a cohesive research paper that we would summarize and present to the class orally. Each week, we would turn in our handwritten, double-spaced notebooks to receive feedback on our writing. We were encouraged to do as much of our research reading in French as possible, and to try to "think in French" rather than writing our entries in English and then translating. For my topic, I chose to write about the influence of Norman French on the English language. This ensured that I would both be working in somewhat familiar conceptual territory, and that I would need to use many words relating to history and linguistics that I was sure I did not already know in French.

This task required me to engage in the acquisition of academic vocabulary in a way that I felt was comparable to the way that my own students would have to when they entered the university—albeit with lower stakes. I had to find words for concepts which I knew in my L1, and also to move target words from my own receptive knowledge domain to the productive domain rather quickly. In order to complete this project, I found I had to utilize word-attack strategies for understanding newly encountered vocabulary, and I found that some of the digital tools recommended by the instructor facilitated that process.

The online vocabulary tools which proved most useful to me during the research and writing were the English-French (and vice-versa) bilingual dictionaries at *Wordreference.com*, and *Linguee.com*, both of which feature an "external sources" section of each entry that locates examples of target words being used in context in a French-English parallel corpus. The French-language version of Google at Google.fr also

proved useful as both a search engine for locating texts and, by using its predictive search feature, a makeshift concordancer.

The online bilingual dictionaries were invaluable for translating the words I knew in English into the needed French equivalents. Often the translation results overwhelmed me with options, though, and the integrated parallel corpus data helped me select the best translation by allowing me to see the word in use in authentic contexts. Usually, the provided context allowed me to identify the translation that was most appropriate for the register in which I was writing. Often this approach allowed me to locate important collocational information as well. For example, I would be able to determine which articles and prepositions were likely to be found in phrases containing the target words, and I could identify lexical chunks that were repeated throughout several samples, which signaled to me that this might be a phrase that I could lift and incorporate into my writing wholesale without fear of plagiarism. The parallel concordancer also helped me avoid the pitfalls of attempting word-for-word translations of multiword items that were idiomatic in English but which either lacked a French equivalent or required a multiword construction that was similarly idiomatic but impossible to guess for a NNS.

Similarly, Google.fr allowed me to locate common collocations and test whether collocations which I suspected to be idiomatic were, in fact, common enough to yield any results. It also provided lexicogrammatical feedback, quickly pointing out when a multiword item I had begun entering did not align with conventional expectations of gender or number agreement. It was also helpful in pointing out misspellings, which is an aspect of vocabulary learning I had rarely thought about from the perspective of a

language teacher, but which now seemed much more prominent from the perspective of a language learner.

Using these digital tools to research vocabulary, I was able to use new words much more accurately than I would have been able to without them. They sped up my writing process and allowed me to move concepts from L1 to L2 with ease. Unfortunately, I found that in some cases I had to repeat searches because I had moved through the process too rapidly, I had not made enough effort to learn the crucial aspects of the word's meaning and usage, and thus I was unable to produce it the next time I needed it. On the other hand, terms that I was required to recycle and produce several times throughout the project became much more automatic by the time I had completed it.

Although I did not use a concordancer and corpus equivalent to the COCA at this stage of my L2 academic writing, I did meet the challenges of the project by taking the initiative and using digital technologies that integrated corpora or concordance-like features. This experience corroborated my theory that concordancing tools can facilitate the process of learning to use new academic and discipline-specific vocabulary.

CHAPTER 5: CONCLUSION

As English learners from around the world begin their studies in English-medium universities, they face the perennial challenge of acquiring the specialized vocabulary of their chosen disciplines. Their competence in using discipline-specific lexis accurately in speech and writing has the potential to mark them as either insiders or outsiders in their chosen academic and professional discourse communities. Mere incidental exposure to these terms in readings and lectures will not necessarily provide the depth of knowledge needed when it comes time for them to produce these lexical items. In order to prepare them to become creative, rhetorically credible participants in their chosen fields, then, intensive English program (IEP) instructors must foreground the study of vocabulary and promote independent vocabulary acquisition strategies in their classes.

Well-planned vocabulary instruction provides a necessary bridge between the receptive (i.e., reading and listening) and productive (i.e., writing and speaking) domains of knowledge. Contextualizing individual lexical items within their most commonly occurring collocational patterns can provide the framework and connective material that will help ELs produce specialized vocabulary to express their ideas clearly and in ways that strike native speakers as "correct." Traditional methods of vocabulary instruction often lack a sound method for teachers and students to systematically access and explore authentic collocational data about new terms. This gap can be addressed by IEP instructors who take the time to become familiar with corpora and concordancing technology in order to share these valuable digital tools with their students.

Thanks to the growth of the internet and the work of many corpus linguists over many decades, IEP instructors and their students now have access to large, specialized corpora which can be mined for samples of authentic language-in-use through increasingly sophisticated concordancing tools. Concurrently with the evolution of these tools, TESL scholars and applied linguists have argued that such valuable resources belong in a learner's vocabulary-learning repertoire alongside more traditional tools like translators and learner dictionaries. The push for student use of corpora and concordancers has only grown, and many research articles have explored the subject from a variety of perspectives.

Many of these studies have found that students' attitudes toward concordancing have been more favorable than critics of student concordancing had anticipated. On the other hand, few of the studies have demonstrated that student vocabulary acquisition, production, and retention have improved through concordancing as much as proponents might have hoped. The research has shown that there are indeed insights to be gained by students through corpus consultation, and that proper training in using the technology is absolutely essential to favorable outcomes. Surprisingly, despite the rise in corpus-based materials being made available for teacher training and curriculum design, resources aimed at promoting independent student concordancing have only appeared sporadically and in limited numbers.

To address this gap in student-oriented concordancing instruction, and to help teachers prepare advanced ELs to meet the lexical demands of English-medium university study, in this M.A. Project I have outlined a unit that scaffolds the process of

independent student corpus consultation. I have provided a set of criteria that teachers can

use to evaluate the ever-growing number of available corpora and to choose the corpus

that best suits their teaching context and aims. I have provided an overview of the COCA,

a large, well-organized corpus that is freely accessible online, and I have identified the

features and search tools of its integrated concordancer that will be most valuable and

least complicated to students who are beginning to explore the technology. I have

provided what I hope will be an easily traversable path for other IEP instructors and their

students to walk through the potentially daunting realm of corpus consultation.

Finally, I have documented my tentative steps in introducing students to

concordance data in my own teaching and in utilizing corpus-based search tools in my

own L2 academic writing. These experiences have made the research and theory I have

been exploring on corpora much more concrete, and they have allowed me to experience

both the power of these tools and some of the challenges that come along with their use.

While independent concordancing may not be the panacea for vocabulary-

acquisition problems that linguists and language instructors might have hoped it to be, it

is nevertheless a valuable tool that can allow students access to rich, authentic data on

language-in-use that is not readily available through any other means. With proper

training and a fair amount of patience, motivated students can integrate corpus

consultation into their repertoire of vocabulary-learning strategies. Concordance lines can

draw students' attention to the ways target items are embedded in multiword items with

varying degrees of frequency, and this information allows for more meaningful

engagement with the terms—and, presumably, promotes better retention as well as more accurate production.

In the future, we are likely to see more student-friendly corpora and concordancing tools become available, and along with these there may be more materials, much like the practical section of this M.A. Project, aimed at training students to use these technologies. As access to these resources becomes greater and more teachers and students become familiar with them, further research will most likely allow applied linguists to determine which tools and methods of instruction in their use are most effective for the enormous variety of teaching and learning contexts.

Ultimately, I hope that the centrality of vocabulary-acquisition strategies to language learning will be recognized by more IEP instructors, and that corpora and concordancers will take their rightful place among the suite of language-learning tools that teachers train their students to use. There is a great deal at stake when it comes to welcoming English learners into academic and professional fields in which English is the lingua franca. For the learners, there are real material, social, and economic consequences which hinge on their ability to establish themselves as accepted members of these communities. For other members of these communities, the insights, ingenuity, and new perspectives that these multilingual students can bring are invaluable. It is my hope that this project has contributed to bringing down the language barrier which impedes the free intermingling of people, cultures, and ideas.

WORKS CITED

Bernardini, Sylvia. "Corpora in the Classroom: an Overview and Some Reflections on
Future Developments" *How to Use Corpora in Language Teaching*, edited by
John Sinclair, John Benjamins, 2004, pp. 15-36.

Biber, Douglas. *Variation across Speech and Writing.* Cambridge UP. 1988.

Boulton, Alex. "Learning Outcomes from Corpus Consultation." *Exploring New Paths in
Language Pedagogy: Lexis and Corpus-Based Language Teaching*, edited by
Maria Moreno Jaen et al., Equinox, 2010, pp. 129-44.

Brun-Mercer, Nicole, and Cheryl Boyd Zimmerman. "Fostering Academic Vocabulary
Use in Writing." *The CATESOL Journal*, vol. 27, no. 1, 2015, pp. 131-48.

Cobb, Tom. *The Compleat Lexical Tutor.* www.lextutor.ca. Accessed 1 October 2016.

---, and Marlise Horst. "Reading Academic English: Carrying Learners Across the
Academic Threshold." *Research Perspectives on English for Academic Purposes*,
edited by John Flowerdew and Matthew Peacock, Cambridge UP, 2001, pp. 315-
29.

Corson, David. *The Lexical Bar*. Pergamon Press, 1985.

Coxhead, Averil. "A New Academic Word List." *TESOL Quarterly* vol. 34, no. 2,
Summer 2000, pp. 213-38.

---. "Academic Vocabulary, Writing and English for Academic Purposes: Perspectives
from Second Language Learners." *RELC Journal*, vol. 43, no. 1, 1 April 2012, pp.
137-45.

---, and Pat Byrd. "Preparing Writing Teachers to Teach the Vocabulary and Grammar of

      Academic Prose." *Journal of Second Language Writing*, vol. 16, no. 3, September

      2007, pp. 129-47.

---, and Paul Nation. "The Specialized Vocabulary of English for Academic Purposes."

      Research Perspectives on English for Academic Purposes, edited by John

      Flowerdew and Matthew Peacock. Cambridge UP, 2001, pp. 252-67.

Cummins, J. "Cognitive/Academic Language Proficiency, the Optimum Age Question

      and Some Other Matters." *Working Papers on Bilingualism*, no. 19, October

      1979, pp. 121-29.

Davies, Mark. *The Corpus of Contemporary American English: 560 Million Words,*

      1990-Present. corpus.byu.edu/coca. Accessed 18 December 2017.

Ferris, Dana. *Response to Student Writing: Implications for Second Language Students.*

      Lawrence Erlbaum Associates, 2003.

Flowerdew, John. "Concordancing in Language Learning." *The Power of CALL*, edited

      by Martha Carswell. Athelstan, 1996, pp. 97-113.

Gabrielatos, Costas. "Corpora and Language Teaching: Just a Fling or Wedding Bells?"

      The Electronic Journal of Teaching English as a Second or Foreign Language,

      vol. 8, no. 4, March 2005, pp. 1-37. Accessed 12 April 2017.

Gardner, Dee. *Exploring Vocabulary: Language in Action*. Routledge, 2013.

---, and Mark Davies. "A New Academic Vocabulary List." *Applied Linguistics*, vol. 35,

      no. 3, 1 July 2014, pp. 305-27.

Gavioli, Laura. *Exploring Corpora for ESP Learning*. John Benjamins Publishing
    Company, 2005.

---. "The Learner as Researcher: Introducing Corpus-Concordancing in the Classroom."
    Learning with Corpora, edited by Guy Aston. Athelstan, 2001, pp. 108-137.

Grabe, William., and Fredricka L. Stoller. *Teaching and Researching Reading*. Second
    ed., Routledge, 2013.

Johns, Tim. "Should You Be Persuaded: Two Samples of Data-Driven Learning." *ELR*
    Journal, no. 4, 1991, pp. 1-16.

---. "Data-driven Learning: the Perpetual Challenge" *Teaching and Learning by Doing
    Corpus Analysis*, edited by Bernhard Kettemann and Georg Marko. Rodopi, 2002,
    pp. 107-17.

Krashen, Stephen. "We Acquire Vocabulary and Spelling Through Reading: Additional
    Evidence for the Input Hypothesis." *The Modern Language Journal*, vol. 73, no.
    4, Winter 1989, pp. 440-64.

Laufer, Batia and Paul Nation. "Vocabulary Size and Use: Lexical Richness in L2
    Written Production." *Applied Linguistics*, vol. 16, no. 3, September 1995, pp.
    307-22.

Laufer, Batia, and Zahava Goldstein. "Testing Vocabulary Knowledge: Size, Strength,
    and Computer Adaptiveness." *Language Learning*, vol. 54, no. 3, September
    2004, pp. 399-436.

Lee, Siok H, and James Muncie. "From Receptive to Productive: Improving ESL

Learners' Use of Vocabulary in a Postreading Composition Task." *TESOL*

Quarterly, vol. 40, no. 2, June 2006, pp. 295-320.

Lewis, Michael. *Teaching Collocation: Further Developments in the Lexical Approach*.

Language Teaching Publications, 2000.

Linguee.com. www.linguee.com. Accessed 30 May 2016.

Lowry, Mary "Lexical Issues in the University ESL Writing Class." *The CATESOL*

Journal, vol. 11, no.1, 1999, pp. 7-37.

Ma, Ka Cheung. *Classroom Concordancing: How and What Students Learn*. 1993. City

Polytechnic of Hong Kong, M.A. Dissertation.

Mikulecky, Beatrice, and Linda Jeffries. *Advanced Reading Power*. Pearson Education,

2007.

Nation, I.S.P. *Learning Vocabulary in Another Language*. Cambridge UP, 2001.

O'Sullivan, Íde, and Angela Chambers. "Learners' Writing Skills in French: Corpus

Consultation and Learner Evaluation." *Journal of Second Language Writing*, vol.

15, no.1, March 2006, pp. 49-68.

Oxford Learner's Dictionaries. www.oxfordlearnersdictionaries.com/us. Accessed 8

April 2017.

Reppen, Randi. *Using Corpora in the Language Classroom*. Cambridge UP, 2010.

Santos, Terry. "Professors' Reactions to the Academic Writing of Nonnative-Speaking

Students." *TESOL Quarterly*, vol. 22, no. 1, March 1988, pp. 69-90.

Schmitt, Diane, and Norbert Schmitt. *Focus on Vocabulary 2: Mastering the Academic Word List.* Pearson Education, 2011.

Sokolik, Maggie. "Digital Technology in Language Teaching." *Teaching English as a Second or Foreign Language*, edited by Marianne Celce-Muria, et al. National Geographic Learning, 2014, pp. 409-21.

Swales, John. "The Concept of Discourse Community." *Genre Analysis: English in Academic and Research Settings*, edited by John Swales. Cambridge UP, 1990, pp. 21-32.

Thinkmap Visual Thesaurus. www.visualthesaurus.com. Accessed 1 October 2016.

Thornbury, Scott. "L is for Lewis." *An A-Z of ELT,* 9 May 2010, scottthornbury.wordpress.com. Accessed 24 April 2016.

Thurston, Jennifer, and Christopher N. Candlin. "Concordancing and the Teaching of the Vocabulary of Academic English." *English for Specific Purposes*, vol. 17, no. 3, 1998, pp. 267–80.

Tongpoon, Angkana. *The Enhancement of EFL Learners' Receptive and Productive Vocabulary Knowledge through Concordance-Based Methods.* 2009. Northern Arizona University, PhD dissertation. search.proquest.com/docview/305067285. Accessed 3 January 2017.

Tribble, Chris, and Glyn Jones. *Concordances in the Classroom: A Resource Guide for Teachers*. Athelstan, 1997.

West, Michael. *A General Service List of English Words.* Longman, Green, 1953.

Woolard, George. "Collocation: Encouraging Learner Independence." *Teaching*

>   *Collocation: Further Developments in the Lexical Approach*, edited by Michael

>   Lewis. Language Teaching Publ., 2000, pp. 28-46.

WordAndPhrase. www.wordandphrase.info. Accessed 20 October 2016.

WordReference.com. www.wordreference.com. Accessed 1 October 2016.

Yoon, Hyunsook, and Alan Hirvela. "ESL Student Attitudes toward Corpus Use in L2

>   Writing." *Journal of Second Language Writing*, vol. 13, no. 4, December 2004,

>   pp. 257–83.

APPENDIX A: LIST OF ABBREVIATIONS

Abbreviation	Meaning
AVL	Academic Vocabulary List
AWL	Academic Word List
CALP	Cognitive Academic Language Proficiency
COCA	Corpus of Contemporary American English
DDL	Data-Driven Learning
EAP	English for Academic Purposes
EL(s)	English Learner(s)
ESL	English as a Second Language
ESP	English for Specific Purposes
IEP	Intensive English Program
KWIC	Key Word in Context
L1(s)	First Language(s)
L2(s)	Second Language(s)
NNS(s)	Non-Native Speaker(s)

APPENDIX B: SAMPLE LEXICAL JOURNAL ENTRY FOR THE TERM
"STRATIFIED"

Term: stratified
Part of Speech: adjective/verb
Definition: arranged in layers or levels
Translation: stratifié
Word Family: stratification, stratify, stratifying, stratifications, stratifies
Collocations: highly stratified, stratified by, stratified analysis, stratified according to
Example: Society in colonial India was highly stratified

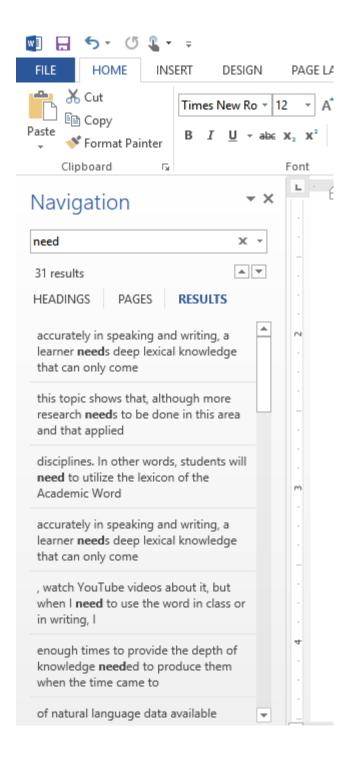APPENDIX C: PRINCIPLES FOR TEACHING ACADEMIC VOCABULARY IN
ENGLISH FOR ACADEMIC PURPOSES

Source	Instructors Should:
Lowry (1999)	1. Promote student awareness of the connection between intentional vocabulary study and effective academic writing
	2. Provide independent vocabulary study strategies
	3. Provide guided practice with new vocabulary
	4. Familiarize students with general academic vocabulary
	5. Provide feedback about students' use of new vocabulary items
Coxhead and Byrd (2007)	6. Encourage students to expand their academic vocabulary
	7. Promote awareness of differences between academic and non-academic vocabulary
	8. Help students understand that memorizing definitions and synonyms is not sufficient
	9. Help students learn academic words by paying attention to common collocates and lexical phrases
Gardner (2013)	10. Help students learn to conceptualize, i.e., associate new terms with their meanings
	11. Encourage form and meaning practice (e.g., flashcards)
	12. Promote context-based word-learning strategies
	13. Train students to use learner dictionaries
	14. Train students to make inferences about new terms based on their prefixes and suffixes
	15. Train students to identify collocational patterns
	16. Encourage students to read texts within the professional fields or academic disciplines they intend to pursue
	17. Encourage students to learn the key vocabulary of the professional fields or academic disciplines they intend to pursue

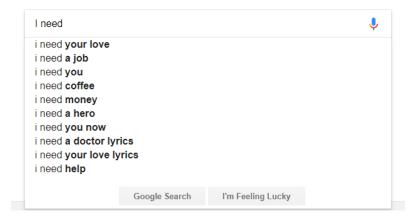| Source	Instructors Should:
Brun–Mercer and Boyd Zimmerman (2015)	18. Teach each vocabulary word's register (e.g., formal, slang, or technical) when teaching the word
	19. Offer a few of each terms' synonyms from other registers to help students distinguish between academic and non-academic words
	20. Teach students how to use learner dictionaries that list the register of each term
	21. Provide text-correction activities in which students find and replace terms from an incorrect register
	22. Provide students with a repertoire of very common academic words and phrases, including logical connectors like "on the other hand"

APPENDIX D: SOME FREE ONLINE ENGLISH CORPORA

Corpus Name and URL	Focus
American National Corpus www.anc.org	15 million words of General American English in 19 genres including court transcripts, jokes, travel guides, and spam
British Academic Written English Corpus www.coventry.ac.uk/bawe	6.5 million words of university-level, proficient, British English student writing in four broad academic genres
Corpus of Contemporary American English https://corpus.byu.edu/coca/	560 million words of spoken English, academic writing, newspapers, and popular magazines
Michigan Corpus of Academic Spoken English https://quod.lib.umich.edu/m/micase/	1.8 million words from transcripts of classroom dialog from the University of Michigan
International Corpus of English http://www.ucl.ac.uk/english-usage/projects/ice.htm	One million words from each of 23 regions where varieties of English are spoken around the world

APPENDIX E: MICROSOFT WORD "FIND" FUNCTION

APPENDIX F: GOOGLE SEARCH PREDICTIVE TEXT

APPENDIX G: WORD CLOUD OF CHAPTER 1 OF THIS M.A. PROJECT