ADDRESSING LANGUAGE ERRORS IN L2 STUDENTS' WRITING:

CAN CORPORA HELP?

A THESIS IN TEACHING ENGLISH TO SPEAKERS OF OTHER LANGUAGES

Presented to the faculty of the American University of Sharjah
College of Arts and Sciences
in partial fulfillment of the requirements for the degree

MASTER OF ARTS

By
SHIREEN PALMER BAGHESTANI
B.A. 2009

Sharjah, U.A.E.
June 2011

We approve the thesis of Shireen Palmer Baghestani   Date of signature

_____   _____
Peter Crompton
Assistant Professor
Thesis Advisor

_____   _____
Cindy Gunn
Associate Professor
 Graduate Committee

_____   _____
David Prescott
Associate Professor
Graduate Committee

_____   _____
Pia Anderson
Program Director, MATESOL

_____   _____
Mark Rush
Dean of the College of Arts and Sciences

_____   _____
Guatam Sen
Office of Research and Graduate Studies

ADDRESSING LANGUAGE ERRORS IN L2 STUDENTS' WRITING:
CAN CORPORA HELP?

Shireen Palmer Baghestani, Candidate for the Master of Arts Degree

American University of Sharjah, 2011

ABSTRACT

Direct corpus consultation by learners has been recognized as a means of helping students speak and write standard English by allowing them to look up real world language usage. However, factors such as students' language proficiency, learning style, and age, in addition to the amount of training students receive in corpus analysis, can impact how effectively they are able to learn from a corpus. The present study therefore aims to determine whether, with minimal training, undergraduate students at the American University of Sharjah are able to benefit from corpus consultation and what attitudes they hold towards it. This study also aims to examine whether students at higher levels benefit more from corpus analysis than students at lower levels. The subjects of the study included two sets of students: freshmen enrolled in a basic writing course, and juniors and seniors enrolled in an advanced writing course. The corpus chosen for the study was *The Corpus of Contemporary American English* (Davies 2008-), the largest, freely available corpus of English on the internet to date.

The findings from this study suggest that students can benefit from direct corpus consultation and that a high proportion of students found the corpus useful. No significant differences were found between Writing 001 and English 207 students in terms of their ability to use the corpus and their attitudes towards it. Therefore, it is recommended that writing teachers at AUS, from the most basic levels to the most advanced, consider adding a corpus component to their classes. Regarding the sufficiency of the training, the results indicate that a large proportion of students would appreciate more training. By the same token, however, the two class periods' worth of training offered was enough to enable students to make many successful corrections to their papers. In order to appeal to students' desire for more training, it is recommended that teachers intersperse corpus activities throughout the semester.

TABLE OF CONTENTS

TABLES

ACKNOWLEDGEMENTS

To my parents. Thank you for instilling in me a love of languages.

توانا بود هر که دانا بود

ز دانش دل پیر برنا بود

(فردوسی)

CHAPTER ONE: INTRODUCTION

The potential for L2 learners to benefit from hands-on corpus consultation has been widely recognized. As such, corpora are no longer considered to be the tool solely of linguists and professional researchers. Students now have free access to large corpora of standard English such as *The Corpus of Contemporary American English* and *The British National Corpus* (Davies, 2008-) through the internet, making them a viable resource for students to consult as a means of ensuring that the language they are using is standard English. By consulting a corpus, students can view actual usages of a word or phrase in professional texts such as newspapers, magazines, and academic journals. Corpora promote autonomy by allowing learners to teach themselves to speak and write standard English. Moreover, it has been found that students who use corpora pay more attention to editing their work in general and feel more confident about their writing ability (Yoon, 2008).

However, it may be that learners are unable to benefit from a corpus as much in practice as they might in theory. As Gavioli (2001) explains, "a concordance may not provide univocal support for a particular analysis and learners, because they are not native speakers of the language, cannot confidently rely on their intuitions to guide and back up their observations" (p. 109). In addition, it has been suggested that without intensive training, students lack the "rigor in observation and reasoning" (p. 81) that is required in order to draw valid conclusions from corpus research (Kennedy and Miceli, 2001).

Furthermore, corpus consultation may prove to be more effective for some students than others. Reading proficiency (Sun, 2003) and language proficiency in general (Yoon, 2008; Liu & Jiang, 2009) seem to be important factors in how well students are able to use corpora. Age and maturity of the students may also play a role, as O'Sullivan and Chambers (2006) found that undergraduates were more overwhelmed by corpora than graduate students.

With these caveats in mind, the present study aims to see whether undergraduate students at the American University of Sharjah are able to benefit from corpus consultation and what attitudes they hold towards it. The corpus chosen for the study was *The Corpus of Contemporary American English* (Davies 2008-) (hereafter COCA). In order to test what effect proficiency, age, and maturity might have on

students' corpus use, the study looked at two sets of students: juniors and seniors enrolled in an advanced writing course and freshmen enrolled in a basic writing course. The students were given only two class sessions of training since, given the full curricula of writing courses, this may be all the training the instructor is able to provide. The following questions, therefore, have been used to guide the research:

1. How successfully do students use a corpus for error correction after two class sessions of training?

2. What are students' attitudes towards corpus use after two class sessions of training?

3. What are the differences between basic and advanced students' proficiency with and attitudes towards corpora?

4. Do students continue to use the corpus even after it is no longer required by the instructor?

5. What types of errors, if any, does corpus consultation help students correct?

By answering these research questions, it is hoped that tertiary level ESL instructors, particularly those at the American University of Sharjah, will have a better idea of whether it is feasible and worthwhile to implement a corpus approach to error correction in their classes. Instructors should be able to gain a better sense of whether corpus consultation is appropriate for the level of students they teach, whether it is reasonable to offer only a small amount of training, and whether their students will actually benefit from using the tool. In addition, by examining the types of errors students have more success with, teachers should have a better idea of how the corpus can help their students.

The next chapter, chapter two, reviews the literature on direct corpus consultation by language learners. Chapter three discusses the methodology used to conduct this study and chapter four presents and discusses the research findings. Chapter five concludes this thesis by answering the research questions and giving suggestions for future research.

CHAPTER TWO: LITERATURE REVIEW

This chapter reviews the literature on applications of corpora in language teaching, focusing specifically on direct corpus consultation by learners. After touching upon the general applications of corpora in language teaching, the chapter proceeds by discussing what direct corpus consultation entails and its theorized benefits for learners. Next, the amount of training and skills needed for students to be able to use a corpus independently are discussed. In addition, the chapter presents the findings of previous studies concerning the extent that students have actually benefitted from corpus consultation. The chapter concludes with recommendations from the literature on how to implement direct corpus consultation in the language classroom.

Corpora in English Language Teaching

Corpora are used in a number of ways to enhance English language teaching. One of these ways is as a source of authentic language data for textbook writers and materials developers. By studying corpora, materials writers are able to obtain a more accurate picture of the way English is spoken in the real world. Another way corpora are used is as a means of studying learner errors. This type of research provides information about the types of learner errors that are most common, for example, in the English of speakers of a particular L1. Because of these two areas of research, it has been argued that corpora "provide a basis for improved pedagogy both by providing better reference tools...and by enabling better decisions as to which lexical items, senses, and grammatical structures should be included in the syllabus" (Aston, 1995, p. 258).

A third way for corpora to be used in English language teaching is as material for the learners themselves. In this case, students directly examine corpus data, usually in the form of concordance lines which have either been printed onto a piece of paper, or which are viewed directly from a computer screen (Gilquin & Granger, 2010). This approach to language teaching is known as data-driven learning or DDL (Johns, 1994; Gilquin & Granger, 2010), and was first introduced over two decades ago. The person credited with the idea, Tim Johns (1986), proposed that corpora can facilitate inductive, form-focused, "discovery" learning. In this approach, students are exposed to authentic, native speaker language from which they make their own discoveries

about the language (Gilquin & Granger, 2010). The belief underlying this approach is that "effective language learning is itself a form of linguistic research" (Johns, 1994, p. 297).

There are many different ways to facilitate data-driven learning, and as long as students are engaging with corpus data, they can be considered to be doing DDL. Data-driven learning activities can range from those which are controlled by the teacher to those in which learners consult the corpus independently (Johns, 1991; Gilquin & Granger, 2010). Another way DDL activities differ is in terms of their presentation. Concordance lines can either be printed onto a worksheet or examined directly from the computer screen. Often, teacher-led activities employ worksheets since they allow the teacher to present selected concordance lines which facilitate the teaching of a particular language point (see, for example, Johns, 1991). In contrast, the more independent the activity, the more likely it is that students will search the corpus directly, using a computer. The next section discusses direct corpus use in more detail.

<div align="center">Direct Corpus Consultation</div>

There are different ways learners can directly consult a corpus. One way is to use it to explore the language in an open-ended manner. For this, students investigate any language area they are unsure about, letting their questions and curiosity drive them (Bernardini 2000a, 2000b, 2004). Students may also use a corpus to discover language to use in their own writing, what Kennedy and Miceli (2010) refer to as "pattern-hunting." This has been shown to work well when students search the corpus for words related to a writing topic, since this often brings up common multi-word expressions associated with the topic or theme.

The present study, however, focuses on a third way of using the corpus, which is as a reference tool. As Chambers (2005) points out, corpora can supplement other reference tools such as dictionaries, grammar books, or the course textbook. Indeed, Johns (1986) believes that this reference function is the "most important...potential use of an interactive concordancer" (p. 161). Just as with other reference tools, students may consult a corpus at different points in the writing process. They may consult it as they are writing, in which case they are using it to answer their own language questions. They may also consult it once the teacher has provided feedback,

in which case the corpus helps them address language errors they did not know they were making at the time they were writing.

As with any reference tool, corpora are thought to enable students to become more independent. The advantage of consulting a corpus, however, over or in addition to other reference tools, is that students can find multiple examples of how a word or phrase is used. This can help students discover different meanings or functions of a word (Liu and Jiang, 2009), as well as how to form correct sentences with the word (Yoon & Hirvela, 2004). Ideally, corpora should be able to help students to produce grammatically-accurate language on their own, or in Aston's (2001) words, "remove much of the need for the teacher to act as an authority concerning the language" (p. 41). With the availability of corpora online, now is a better time than ever to introduce students to corpora since they can access them from home or anywhere they have internet access. In fact, Gilquin and Granger (2010, p. 364) hope to one day see corpora used by learners as frequently and as naturally as dictionaries.

Benefits of Corpus Consultation

Corpus consultation has been shown to have many positive effects on learners. In addition to promoting learner autonomy, Yoon (2008) found that having access to a corpus made students more likely to include an editing step in their writing process. One of the things that makes corpora appealing to students is how quickly they provide answers. Chambers and O'Sullivan (2004) found that students preferred using a corpus over a dictionary because the corpus was faster. Another positive effect Yoon (2008) and Yoon and Hirvela (2004) discovered was that students' confidence increased as a result of consulting a corpus. This is likely due to the fact that it offered students a way to test their language hypotheses and make sure their language use was acceptable (Yoon and Hirvela, 2004).

Gilmore (2008) also points out that having students correct their own errors ultimately benefits them more than if the teacher were to simple tell students the answers. When students use a corpus to correct their errors, they must utilize inductive reasoning by examining language data in order to draw their own conclusions. This type of work is thought to facilitate students' active noticing of linguistic forms (Aston, 1995), which is partly why inductive or "discovery" learning is thought to lead to better retention of what is learned (Liu & Jiang, 2009). Breyer (2009) also remarks that this inductive approach allows students to "discover

language use at their own pace" (p. 163). In addition, inductive learning is thought to be very personal, as students must draw on their existing knowledge which ensures that they will learn what they are ready to learn, an idea linked to constructivist theories of learning (Gilmore, 2008). As Gilmore explains,

> With each learner's interlanguage system in its own unique stage of development, inductive approaches, which encourage students to find their own solutions to their own particular problems, are likely to create the conditions necessary for language acquisition to occur. Of course, as classroom activities, they are also more time consuming, but the increased cognitive work they require should also lead to greater learning gains. (Gilmore, 2008, pp. 365-366)

In addition to allowing students to learn at their own pace, corpus-based instruction can help to foster a critical understanding of grammar (Liu & Jiang 2009; Gavioli & Aston, 2001). In other words, examining corpora gives learners the opportunity to question the validity of the prescriptive rules offered in textbooks and reference materials and empowers students to "establi[sh] their own views of language reality" (Gavioli & Aston, p. 243). Encouraging students to adopt a critical perspective also helps them to realize that native speakers do not always follow a set of rigid grammar rules (Liu & Jiang, 2009), which can teach students to adopt an open mind to language variation among speakers. It may also help train students to have a flexible mindset when encountering language that does not conform to the rules stated in grammar books.

The particular language errors a corpus can help students with are wide and varied. Many of the errors a corpus can be useful for can be classified under lexicogrammar (Liu & Jiang, 2009). Students have reported that using a corpus helps them produce idiomatic expressions and allow them to discover "prototype" phrases or "templates" that native speakers use (Chambers & O'Sullivan, 2004; Kennedy & Miceli, 2010). Likewise, Wu, Witten, and Franken (2010) state that corpora are "a particularly productive context in which to study collocations" (pp. 83-84). Corpora can also elucidate subtle grammatical nuances (Wu, Witten, and Franken, 2010, p. 100; see also Liu & Jiang, 2009; Johns, 1994; Yoon, 2008), such as the difference between two similar lexical items such as *"high"* and *"tall"* (Tsui, 2004) or *"convince"* and *"persuade"* (Johns, 1991). Finally, a corpus can also be useful for

non-lexicogrammatical issues such as capitalization and finding alternatives for overused words (Chambers & O'Sullivan, 2004).

How Much Training Do Students Need?

The general consensus among researchers is that the more training students receive, the better equipped they will be to use a corpus. In fact, in most of the studies published to date, no matter how much training students receive there seems to be a desire among either the students or the researcher(s) for more training. For instance, O'Sullivan and Chambers (2006) found that when asked to comment on factors that would make them continue using corpora, students responded "More assistance and more training" (p. 63). Likewise, students in Chambers & O'Sullivan's (2004) study expressed a need or desire for more training, particularly "more hands-on time" (p. 169) and more explanation of what a corpus can be used for. From the standpoint of researchers, Gaskell and Cobb (2004), who offered four weeks of training through precasts (described below) as well as in-class instruction, still believed that the training offered was "probably insufficient" (p. 315).

In addition, studies which have found students' corpus research skills to be lacking usually cite insufficient training as one of the main factors. For instance, Kennedy and Miceli (2001), who offered an enormous thirty-three hours of training, concluded after examining their students' corpus use that "our training had not adequately equipped them as 'corpus researchers'" (p. 81). Liu and Jiang (2009) echo the need for more training since their students complained of not being able to analyze concordance data effectively. In addition, some have made the case for more training by pointing out that students who receive more training tend to have a more positive experience with concordancing. For instance, Yoon and Hirvela (2006) found that out of two classes featured in their study, the class which had not used corpora as much throughout the semester had less favorable attitudes towards concordancing.

While training is clearly important, Boulton (2009) points out that students can still benefit from concordancing even with limited corpus skills. In addition, Kennedy and Miceli (2001), despite saying their training had not fully prepared their students as corpus researchers, still acknowledge that there was much their students *could* do. For instance, they say "We found that the students made many successful investigations, demonstrating a general appreciation of the types of questions that can be posed, a certain ability to work by analogy, and a preparedness to review their

7

strategies when a search was leading nowhere" (p. 81). Likewise, Chambers and O'Sullivan (2004) found that students, despite being fairly new to corpus research, made a high number of positive corrections to their essays using a corpus.

Despite the fact that more training is usually better, teachers who wish to implement corpus use in their classrooms face practical constraints such as the amount of time they can devote to training. Teachers must, therefore, decide how much training they feel is enough given what they want students to be able to do with the corpus. There are a number of studies based on real language classrooms, and in each the teacher offered a different amount as well as different delivery of training. The fact that training differs so much from study to study shows that there is no standardized, or "right" way of conducting corpus training.

The training offered in the following studies was either short and compact, where it lasted only a couple weeks or class sessions, or was ongoing throughout a whole semester and integrated with regular class activities. Examples of the short and compact approach can be found in three studies which were all conducted at the University of Limerick (Chambers, 2005; Chamber & O'Sullivan, 2004; O'Sullivan & Chambers, 2006). Students were enrolled in a course on how to use technology in language learning. The course contained a three week corpus module during which students received their training.

Another example of the short and compact approach can be found in Gaskell and Cobb's (2004) study. In this study, the researchers provided training in the form of in-class demonstrations, but also through in-text feedback. This feedback was provided in the form of "precasts," or links to an online concordance that the teacher had selected specifically for each student. The precasts served as a way of getting students to practice inducing information from concordances without them having to go through the step of finding the concordance. The use of precasts was discontinued after four weeks, after which students were responsible for finding their own concordances.

There are also a few studies in which students were trained in just one sitting. For instance, both Wu, Witten, and Franken (2010) and Turnbull and Burston (1998) trained students in one session lasting about two hours. The session included demonstrations on how to search the corpus and selecting appropriate search terms as well as tips for analyzing concordance data. They also supplied learners with a user

guide or copy of general guidelines. Similarly, Sun (2003) who only had three students to train, gave them each an individual lesson lasting thirty minutes. Finally, Gilmore (2008) gave his students a thirty minute demonstration and then one hour in the computer lab of free exploration.

Other studies employed training that was integrated into the course and lasted for the duration of the semester. Kennedy and Miceli (2001) used what they refer to as a "gradual and guided approach" (p. 79) in which students were slowly introduced to concordance work starting with controlled, teacher-led activities. The training also included demonstrations of how to use a corpus in revising a written text and later on, students undertook independent corpus investigations. The total amount of training provided over the semester was thirty-three hours.

A similar approach to training was taken by the instructor whose classes were investigated by Yoon and Hirvela (2004) and Yoon (2008). This teacher "incorporated the corpus approach into the curriculum as part of the regular classroom activities" (Yoon, 2008, p. 33). As a guided training activity, he had students construct "prototype strings," or examples of a word's usage patterns, first in pairs and then later on their own. Following this, the instructor moved students into independent problem-solving by having them use the corpus to correct their own errors. He required them to e-mail him their search results weekly, and then would combine their results into a handout given to everyone. The instructor usually devoted twenty minutes per week to discussion of students' corpus work.

Skills Needed for Effective Use of Corpora

The two basic skills needed in order to perform corpus research are the ability to form queries and the ability to analyze concordance lines (Sripicharn, 2010). Being able to formulate appropriate searches increases the likelihood that students will retrieve "relevant and organised outputs that facilitat[e] searching, analysing and making inferences" (Sun, 2003, p. 611). While it is possible just to type in a word and press "search," there are more sophisticated ways of searching which may yield better results, such as using wildcards (where the system searches for any word, word part, or symbol) and lemmas (putting the word in brackets to search for all possible word forms). Students should also know how to select appropriate sub-corpora to search given the type of writing they are doing, as well as be able to improvise by changing

the search terms if an investigation reaches a dead end (O'Sullivan & Chambers, 2004).

The ability to form appropriate searches also depends on the extent to which students understand the problem they are trying to solve (Sun, 2003; O'Sullivan & Chambers, 2006; Chambers & O'Sullivan, 2004). If, for example, the teacher has underlined an entire phrase to indicate there is an error, the student may have no idea what the error actually is, preventing them from forming any kind of relevant search. It may also be that the student has only a vague understanding of the problem and therefore is unable to articulate it and put it into search terms. It is therefore unrealistic to expect students to benefit from a corpus when their understanding of the problem at hand is so limited (Sun, 2003).

However, it is not enough to simply find a concordance that relates to the error, as students also need to be able to correctly analyze the data. This requires inductive reasoning abilities since students must extract patterns from the data. Watson-Todd (2001) has pointed out that the inductive process of making meaning of a concordance is not very different from inductive teaching methods in general. The crucial difference, he explains, is that usually in inductive teaching methods, the teacher selects examples that make the language point very obvious whereas the examples a corpus gives have not been hand-picked to make induction of the rules easier. This points to the fact that students need to be prepared to deal with possibly irrelevant or confusing data when trying to find answers in a corpus.

Not to be overlooked, another problem Watson-Todd (2001) points out is that once students have induced patterns, one should not take for granted that they will be able to apply them towards resolving the error. This step is yet another area where students' corpus investigations can possibly go awry. Finally, a very obvious skill students must possess is fluent reading ability. As Sun (2003) points out, "poor reading proficiency can also limit the learning outcome due to the need for massive reading of concordancer data" (p. 611).

How Successful Are Students at Using Corpora?

Since the ability to infer patterns is crucial to being able to learn from corpora, Watson-Todd (2001) conducted a study in which one of the aims was to discover whether students are able to induce valid patterns from concordance data. The subjects used in his study were of lower intermediate to intermediate level, so they

were not overly advanced in English. Moreover, they received no training in inductive reasoning, apart from being given "an example concordance with induced patterns to follow as a model" (p. 95). Watson-Todd found that the majority of patterns students induced were usually accurate reflections of what was in the concordance. The suggests that even lower-proficiency learners are capable of inducing patterns from corpus data. Moreover, Watson-Todd (2001) found that in twenty out of twenty-three instances, students were able to apply the patterns they had induced to solving their writing problems.

There have also been a number of studies which measured the number of writing errors students are able to resolve using a corpus. Most studies have found students to have a fairly high success rate. For instance, as part of the aforementioned study, Watson-Todd (2001) found that eighteen out of twenty-three students (78%) made successful corrections to their writing using information they had induced from a corpus. In addition, Gilmore (2008), whose subjects were forty-five intermediate Japanese learners of English, found that students were able to improve 61% of errors after they consulted a corpus. Likewise, O'Sullivan and Chambers (2006) found that with their group of undergraduate learners of French, students successfully resolved nineteen out of twenty-five preposition errors (76%) and twenty-two out of twenty-eight word choice errors (79%) using a corpus.

Another study which examined error correction, Wu, Witten, and Franken (2010), found that out of ninety-five attempted corrections, students successfully resolved seventy-three of them (77%). Subjects in this study were students in an IELTS writing preparation class in New Zealand. The researchers found, however, that the success rate for correcting lexical collocation errors was higher (70.5%) than the success rate for correcting grammatical errors such as verb morphology (55%). In addition, Gaskell and Cobb (2004), who studied lower intermediate Chinese learners of English in Montreal, found that students made successful corrections between eighty and one hundred percent of the time when given precasts, and between sixty and seventy percent of the time without the help of precasts.

Not all studies have such optimistic findings about students' corpus research skills, however. For instance, Kennedy and Miceli (2001) provide an in-depth account of specifically where students' research skills were lacking. The students whom they based their analysis on were eight learners of Italian whom they videotaped while

revising an essay using a corpus. The first deficiency they point out was that students had difficulty forming the research question. They often failed to ask the "right" question in that it did not properly address the error students were trying to correct.

Second, Kennedy and Miceli (2001) found that students lacked sophistication in the way they searched the corpus. For instance, they frequently failed to realize that they could search for different inflected forms of the keyword in the event that the first form they tried retrieved too few results (see also O'Sullivan & Chambers, 2006). Students also did not make good use of the more advanced search functions, such as the option to specify whether to allow intervening words between the keywords. Third, when students analyzed the concordance lines, they fell into certain "traps" (p. 86) that prevented them from locating the lines that would be of help to them. For instance, they would be so fixated on a certain word that it caused them to overlook the useful examples that were right in front of them. As an example, students were so bent on finding which preposition to use with a certain verb that they did not see the lines which illustrated that that particular verb did not actually need a preposition. Finally, students' reasoning was often flawed: for example, if they failed to locate a particular item in the corpus, their conclusion was "this does not exist in the language" rather than "there is no example of this in the corpus."

O'Sullivan and Chambers (2006) reiterate many of the difficulties cited by Kennedy and Miceli (2001). Their analysis was also based on students using a corpus to revise their essays, but rather than videotaping them, the researchers had the students fill in feedback forms describing the searches they had carried out. Like Kennedy and Miceli, the authors noticed that students often forgot to try variants of the search word when the initial search failed. Students also had difficulty forming the research question, as they "failed to look beyond the underlined word or phrase to try to resolve the problem" (p. 60). Moreover, like the previous authors they found that at times students "misinterpreted the outcome of the searches" (p. 60).

What Factors Determine Students' Success with Corpora?

Many variables can affect a student's experience using a corpus. One of these variables is the student's understanding of the error they want to resolve. Sun (2003) refers to this understanding as "prior knowledge." She explains that the learner's prior knowledge will either be complete, incomplete, or nearly zero. For students with complete prior knowledge, the corpus may simply be a way "to confirm a hunch"

(O'Sullivan and Chamber's, 2004, p. 58). However, when students' prior knowledge of a grammatical point is nearly zero, as when the teacher has underlined something but the student has no idea what is wrong with it or how to fix it, the corpus may be of little use (Gilmore, 2008). One student in Chambers and O'Sullivan's (2004) study corroborated this notion by saying "You have to have a clear idea of what alternatives to look for so your own subject knowledge has to be pretty advanced."

The need for a certain amount of prior knowledge in order to benefit from corpus consultation echoes Hunston's (2002) recommendation that corpora are best-suited for "very advanced learners who are filling in gaps in their knowledge rather than laying down the foundations" (p. 171, as cited in Gilquin & Granger, 2010, p. 363). For instance, students may use a corpus as a tool for double-checking that what they have written is correct, or for looking something up they are unsure of, such as a preposition.

Closely related to prior knowledge is the issue of language proficiency. Clearly, students with higher language proficiency are likely to have an easier time using a corpus. Low language proficiency may be one reason why certain students foster negative attitudes towards corpus use (Liu & Jiang, 2009; Yoon, 2008). In addition, students may find the number of examples generated by the concordancer overwhelming (O'Sullivan & Chambers, 2006; Liu & Jiang, 2009). This is why, as Sun (2003) points out, reading proficiency is so important to successful use of corpora.

Learning style is another factor that can determine students' success with corpora (Gilmore, 2008; Turnbull & Burston, 1998). As with all teaching approaches, corpus consultation may appeal to some students more than others. While some may really appreciate having the corpus to assist them, others might view concordancing as just another school assignment. As a case in point, Gaskell and Cobb (2004) required students to use corpora for the first four assignments, but after that corpus use was optional. Out of twenty students, seven of them continued to voluntarily use the corpus after it stopped it being required. Likewise, Turnbull and Burston (1998) in their case study found that while one student experienced great success with concordancing, the other did not--suggesting that certain students will have more of a natural ability for corpus analysis than others.

Although natural ability likely plays a role, Kennedy and Miceli (2001) also believe that certain corpus analysis skills are not intuitively obvious and must therefore be taught. This includes teaching students how to use a concordancer, formulate searches, and analyze the data once it has been retrieved. As part of analyzing the data, students must possess inductive reasoning skills. As noted by Watson-Todd (2001), "Concordances can highlight grammatical patterns, collocations and pragmatic aspects of lexical items...but these need to be induced from the corpus" (p. 93). In other words, a corpus doesn't provide students with the answers, but supplies them with the data needed to find the answers. Thus if students are inexperienced with inductive learning strategies, learning from corpora may be a struggle (Sun, 2003).

Age and maturity level may be another factor to consider when introducing students to corpus analysis. According to O'Sullivan and Chambers (2006), their postgraduate students displayed more positive attitudes towards corpus use than their undergraduate students. In addition, Liu and Jiang (2009) noticed a correlation between motivation and positive attitudes towards corpora--that is, students who were more motivated tended to get more out of corpus activities. This may have been due to the fact that analyzing a concordance is an active process and students who are unmotivated might not be willing to make the effort. On the other hand, it is important to bear in mind that even when motivation levels are equal, some students may simply enjoy inductive learning more than others (Liu & Jiang, 2009).

Students' Attitudes Towards Corpus Use

In any given study, it is interesting to look at the proportion of students who held positive attitudes towards corpus use. In most cases, the proportion of students with positive attitudes towards corpus use was quite high. For instance, Gilmore (2008), who surveyed students after just using the corpus once, found that 95% believed that "online corpora were a useful resource to aid them in redrafting their essays" (p. 367). In addition, O'Sullivan and Chambers (2006) found that ten out of the fourteen students they surveyed (71%) reported that they had found concordancing helpful in improving their French writing skills. Also, Chambers and O'Sullivan (2004) report that all eight of their participants had found corpus consultation helpful.

Another study which found a high positive response from students was Yoon's (2008), in which participants "praised the advantages of corpus use in L2 writing" (p.

43). Yoon and Hirvela (2004) also reported that some students liked to have the corpus open any time they were writing, whether it was a paper for school or just an email. They also interviewed a student who commented that he thought "all ESL writing courses should add a corpus component" (p. 274).

Other studies reported more mixed attitudes towards corpus use. For instance, Gaskell and Cobb (2004) found that out of twenty students, only eight (40%) attributed their improvement in grammar to corpus work. Liu and Jiang (2009) also found that only 39% responded that they had learned either "a great deal" or "a good amount" from the use of corpora. At the same time, 49% said they had only learned "a little." However, they found that the students with positive attitudes towards corpus consultation "hail[ed] the use of corpora as an excellent weapon for conquering their language learning problems" (p. 75).

Another revealing finding is the number of students who take up voluntary use of corpora after being introduced to them. As already mentioned, Gaskell and Cobb (2004) found that seven students out of twenty (35%) continued to consult the corpus after it was no longer required. In addition, O'Sullivan and Chambers (2006) asked students whether they thought they would continue to use corpora and six (46%) answered "yes," four (31%) answered "maybe," and three (23%) answered "no." Liu and Jiang (2009) asked the same question of participants in their study and found that 52% of students believed they would continue to consult corpora. The general trend these studies indicate is that somewhere around fifty percent of students in any given course tend to continue using corpora or believe they will continue to use them.
Recommendations for Doing Concordancing Work with Students

When carrying out concordancing work with students, particularly when teaching students how to consult a corpus, there are a number of considerations to bear in mind. One of these is that students' attitudes towards concordancing may change as they become more familiar with the tool. In particular, Yoon (2008) and Yoon and Hirvela (2004) found that students' attitudes became more positive the more they used the corpus. Similarly, Gaskell and Cobb (2004) noticed an increase in voluntary corpus usage towards the end of the course, attributing this to students' increased confidence with the technology.

The teacher's presence also plays an important role in helping students acquire the skills necessary to use corpora. Having a smaller class size is clearly beneficial, as

this means the teacher can devote individual attention to students more frequently (Liu and Jiang, 2009). It is also helpful to use a "gradual and guided approach" (Kennedy & Miceli, 2001, p. 79), where students receive instruction over a long period of time, such as over the course of a semester (Kennedy & Miceli, 2001; Yoon & Hirvela, 2004; Yoon, 2008), and where the teacher provides numerous opportunities for practice. Yoon and Hirvela (2004) found that students were more likely to enjoy using corpora when they "played a greater role in terms of in-class activity" (p. 269).

There are numerous corpus-based activities for teachers to choose from. For instance, teachers can begin by going over paper-based concordances with students rather than having them start by searching the corpus independently. Also, once students begin searching the corpus independently, teachers can have them discuss their findings with the class. These discussions give students an opportunity to practice using the language, as well as learn from each other's finding (Gavioli & Aston, 2001; Liu & Jiang, 2009). One of the caveats, however, is that learners may lack the grammatical metalanguage to articulate what they find (Breyer, 2009), although Gavioli and Aston (2001) believe that this process allows students to "develop their own descriptive frameworks" (p. 242).

Another suggestion, whether students are analyzing a paper-based concordance or searching the corpus on their own, is to have them start with deductive exercises before moving onto inductive exercises (Liu & Jiang, 2009). Deductive exercises are generally seen as being easier, since they require students to search for evidence exemplifying or contradicting rules or patterns they have already learned (Liu & Jiang, 2009). Since deductive exercises are easier, they can serve to familiarize students with the basic elements of concordances so that they have a groundwork for when they begin to do inductive work.

Conducting corpus searches with a partner or small group is another recommended way to ease students into corpus analysis. Liu and Jiang (2009) found that students who were new to using a corpus liked working in groups since they could collectively figure out how to search the corpus. In addition, concordances usually yield a high number of tokens, which can be overwhelming for one student to analyze on their own. Working in a group allows the students to divide up the work.

Also, because inferring rules and patterns is a difficult task, group members can help one another by contributing their own points of view (Liu & Jiang, 2009).

Another option for teachers to consider is simplifying the corpus itself. Gavioli and Aston (2001) recommend using small, homogeneous corpora with students in the beginning. The benefit of using a smaller corpus is that students will not be as overwhelmed by the number of examples. Likewise, a homogeneous corpus, one consisting of highly similar texts, makes the data easier to analyze since there will be "a substantial number of recurrent features" (Gavioli, 2001, p. 118). However, small, homogeneous corpora will likely need to be created by the teacher, which may not be practical for all teachers.

One last technique for training students to use corpora is for the teacher to model how to conduct searches (Sun, 2003; Liu & Jiang, 2009). As already mentioned, devising effective searches is one of the necessary skills for being able to use a corpus. Turnbull and Burston (1998) believe that students greatly benefit from "models on which to base the formulation of appropriate search items" (p. 18). In addition, students should have the opportunity to practice devising searches with the teacher available to provide feedback and assistance (Sun, 2003).

To summarize, research suggests that corpora are rich with possibility for enhancing students' language awareness and empowering them to be autonomous learners. Corpora have been shown to facilitate language learning and have received high approval ratings from many learners. Still, teachers should bear in mind that some students are more positively predisposed to corpus consultation than others, and should not therefore expect all students to benefit to the same degree. At the same time, most would agree that training should be provided in order for students to gain the most from corpus consultation. Teachers can certainly benefit from more research into how to effectively implement corpus work with students.

CHAPTER THREE: METHODOLOGY

This study aims to see whether and to what extent tertiary ESL students can benefit from corpus analysis with limited training. The present chapter details the methodology used to carry out the study. First, the participants and setting of the study are described, followed by the methods of data collection. These included an empirical study, two surveys, and student interviews. The final section contains a description of the methods used to analyze the data.

Participants and Setting

This study was carried out with 99 students at the American University of Sharjah enrolled in two different writing courses. Thirty-four of the students were enrolled in Writing 001, a "fundamentals" writing course and the most basic writing course offered at the university (henceforth referred to as WRI). The remaining 65 students who participated in the study were enrolled in English 207, the culminating writing course for engineering majors (henceforth referred to as ENG). Three sections of each course, WRI and ENG, were included in the study, for a total of six sections. The two courses were taught by different instructors, but both instructors taught all three sections of their respective courses. Both courses, WRI and ENG, met three times a week for 50 minutes each time.

The students who participated in the study came from a variety of language backgrounds. According to the 78 participants who provided demographic information, seventy percent were native Arabic speakers, ten percent were native Urdu speakers, and eleven percent spoke Farsi, Malayalam, Italian, Turkish, or another language as their first language. Nine percent either spoke English as their first language or indicated that they were as or more proficient in English than their first language. Therefore, 9% of the participants can be considered native or near-native speakers of English, while 91% would be considered non-native speakers. However, 21 participants are unrepresented in these figures as they did not provide demographic information.

Methods of Data Collection

There were four types of data collected for this study. The first was an empirical study designed to assess how well students were able to use a corpus having received minimal training. The second was an attitude survey administered to students directly

after completion of the empirical study designed to assess students' attitudes towards corpus use. The third data collection method was interviews with seven students to follow up on what they wrote in their surveys and to obtain more in-depth feedback. The fourth was a follow-up survey administered six weeks after the empirical study to see whether students continued using the corpus.

Empirical Study

The empirical study consisted of training the students to use COCA and having them do an in-class error correction assignment using COCA once they had been trained. This error correction assignment was the instrument used to measure how well students were able to use the corpus. The empirical study was conducted over the period of two weeks. It was conducted with the WRI students the first week, and with the ENG students the second week. Each class therefore participated in a week, or three 50-minute class periods, of corpus training and activities.

Training

The WRI students were trained for two full class periods which left one class period for them to do their error correction assignment. The ENG students, on the other hand, received only one and a half class periods of training since they seemed to catch on faster than the WRI students. This left one class period for the WRI students to do their error correction assignment, and one and a half class periods for ENG students to do their error correction assignment.

The training approach taken in this study was short and compact. Students were introduced to the basics of using corpora such as what a corpus was and various search techniques which could be used. This approach was similar to those taken in other studies (Turnbull and Burston, 1998; Gilmore, 2008; Wu, Witten, and Franken, 2010). Despite numerous recommendations in the literature to train students gradually and offer them many chances to interact with the corpus during class, the short and compact approach was used in order to see what benefits, if any, students could gain having only received minimal training.

The list of topics covered for both classes (see Table 1) was the same except WRI students were not taught how to use the collocation search feature of COCA. This was a conscious decision so as not to overwhelm the WRI students with too many search techniques. The WRI students also received a worksheet which had questions designed to help them better understand what COCA was and which provided some

searches for them to do as practice (see Appendix A). The ENG students did not receive this worksheet, but were instead provided with a handout posted on Blackboard, describing the different types of searches (see Appendix B). However, it seems as though a number of ENG students did not retrieve the handout from Blackboard, as several of them commented in their surveys that they wished a handout had been provided.

Table 1: List of Topics Covered in Training Sessions

- What is a corpus?
- What is COCA?
- Why use a corpus instead of a dictionary?
- Searching for one or more words (basic search)
- Seeing expanded context
- Filtering by text type
- Jumbling the results
- Using the COLLOCATES search field
- Searches with punctuation
- Lemma search
- Specifying the part of speech of a word
- Wildcard attached to the word (e.g., work*)
- Wildcard with specified part of speech (e.g., resulted [i*])
- Wildcard search (e.g., resulted *)
- Using the KWIC display

Error Correction Assignment

As mentioned earlier, the WRI students had one 50-minute class period to work on their error correction assignments. The majority of them finished early, but some did use the full 50 minutes. The ENG students had one and a half class periods to do their error correction assignments since their training ended early in the second class. All of the students in ENG used the remainder of class on day two and part of class on day three to complete their assignments. None of them, however, needed the full class period on day three to finish their assignments.

The error correction assignment had students use the corpus to correct their own writing assignments. The assignment WRI students worked on correcting was a three-page autobiography. The assignment ENG students worked on correcting was a two to four page paper discussing the attributes of a team and the qualities engineers should possess in order to work effectively in teams. As in the research of Kennedy and Miceli (2001), the errors in each essay were underlined, but no further explanation was provided. This meant that students were responsible for figuring out the nature of the error. The errors that were underlined were those involving capitalization, spelling, form, and usage. Run-ons, fragments, and comma splices were not underlined since it was not expected that students would be able to resolve them using a corpus.

In addition, a week before the corpus training the WRI students received their assignments back, with the errors underlined, and were asked to spend the class period correcting their mistakes. This was done in order to cut down on the number of careless errors in students' papers prior to them doing the error correction assignment with the corpus. It was thought that by eliminating or reducing careless errors, students would be more likely to use the corpus for correcting errors they would otherwise be unable to correct on their own. This step was not taken with the ENG students due to time constraints.

For the "real" error correction assignment, students received their papers back, with the errors underlined, and were asked to correct two errors of their choosing. For the WRI students, the errors that students had failed to correct the first time were circled with a different color pen and a "C" for "corpus" was written beside the error. The reason for allowing students to choose which errors to correct was because, as Sun (2003) points out, a corpus is more likely to help the student if s/he has some idea why the error might be wrong. Giving students the ability to choose which errors to correct was a way of getting them to correct errors they had a better grasp on.

In order to document the students' error correction process, they were asked to fill in an error correction worksheet for each error they corrected using COCA (see Appendix C). This worksheet was adapted from Liu and Jiang (2009). It asked students to write down the original error, the correction they made, and the reason they made that correction (as in Gaskell & Cobb, 2004; O'Sullivan & Chambers,

2006; Liu & Jiang, 2009). The purpose of the worksheet was to provide a record of what the students found in COCA and whether or not their correction was successful.

It was decided that students would not be graded on the number of successful corrections they made, since this might cause students to tackle "easy" errors that they really did not need the corpus for and would invalidate the results since students would not be using the corpus out of genuine need. Therefore, students were only graded on participation--whether or not they filled out a worksheet for each error they corrected. It is unclear whether students put their full effort into doing the assignment, but it is likely that many did not since it was only for a participation grade.

Attitude Survey

A survey consisting of eight questions (see Appendix D), roughly adapted from Yoon and Hirvela (2004), was administered to students via an online survey tool on the same day that they completed their error correction worksheets. The survey was designed to collect information on students' prior experience with corpora, as well as elicit students' attitudes towards corpus use, including whether they believed it was useful and whether they felt they understood how to use it. Thirty-two students from WRI took the survey, but one student failed to answer any of the questions. Therefore, only 31 WRI surveys were recorded and analyzed. Sixty-two surveys were collected from the ENG students.

The surveys administered to the WRI and ENG students were the same except for one question (question four) which was changed after surveying the WRI students and prior to surveying the ENG students. The question was changed because it was thought that the new question would elicit more interesting data than the one on the WRI survey. Question four on the WRI survey asked students to indicate their level of agreement with the statement, "I think the corpus is a useful tool for students wanting to improve their English." The new question, which was included on the ENG survey read, "The corpus provides answers to my questions" with the options "Most of the time," "Sometimes," or "Rarely." Item four of the survey obviously cannot be used to compare the attitudes of WRI students and ENG students, and must be treated as only revealing of the attitudes of the group of which the question was asked. However, as the rest of the survey questions were the same in both surveys, those questions can be used to compare the attitudes of both groups.

Interviews

Three students were interviewed from WRI and four were interviewed from ENG. The selection process was based simply on which students responded to the researcher's email. From the WRI students only three students responded, and all of them happened to be boys. For the ENG students, one male student volunteered himself in person. The other male student and the two female students were chosen as they were the first ones to respond to the email.

All of the interviews took place within a week after the students taking the survey. This was done in order to ensure that the experience of using the corpus was still fresh in the students' minds. The interviews consisted of questions which asked students to elaborate on their survey responses, as well as new questions that were written specifically for the interview. For instance, students were asked if they used other tools to check their writing, and if they felt one needed to have a good understanding of grammar to be able to use the corpus.

Follow-Up Survey

The follow-up survey (see Appendix E) was administered in the tenth week of the semester, which was six weeks after the WRI students had received their introduction to COCA and five weeks after the ENG students had received theirs. The purpose of the follow-up survey was to find out whether students chose to continue using the corpus after it was no longer required by the teacher and the reasons behind their decision. In addition, the survey asked students whether they would like to receive more training. Asking this question after students had been given five to six weeks to use the corpus, as opposed to days after they were first introduced to it, was thought to have allowed students a chance to provide a more informed answer about whether they really needed more training or not. The follow-up survey was also used as an opportunity to collect demographic data from the participants about whether they were native or non-native speakers of English.

Methods of Data Analysis

Empirical Study

The error correction worksheets were analyzed to see how many successful versus unsuccessful corrections there were, as well which types of errors tended to be successfully corrected with the corpus and which tended to be unsuccessfully

corrected. The ratio of successful to unsuccessful corrections was also compared between the WRI and the ENG classes.

Attitude Survey

The survey consisted mainly of closed-ended questions such as yes/no and scaled (Likert-type) questions. There was one open-ended question which invited students to provide additional comments on their experience. The closed-ended questions were analyzed according to the number of students who selected each answer. The open-ended question was analyzed by grouping answers thematically. For instance, some comments provided suggestions on how to improve the training, while others discussed why or why not they felt the corpus was helpful.

Interviews

The interview responses were analyzed in two phases: where the same question was asked of multiple respondents, the data was organized by question so that the different answers could be compared. Where different respondents provided similar ideas or sentiments, though not necessarily in response to the same question, their responses were organized thematically. For instance, more than one person commented that the design of the corpus interface made it easy to use, though this sentiment was not necessarily triggered by the same question. In addition, conflicting sentiments were also noted, as these demonstrated how two students could have different experiences with the same tool.

Follow-Up Survey

The follow-up survey consisted of both closed- and open-ended questions. The closed-ended questions were analyzed according to the number of students who selected each answer. The open-ended questions, in which students had to provide reasons for their continued use or lack of continued use of the corpus, were analyzed thematically to determine the various reasons students did or did not continue to use the corpus.

CHAPTER FOUR: ANALYSIS

This chapter provides an analysis of the data collected for this research. The first half of the chapter presents the findings of the research in the following order: first, the results of the empirical study are presented, followed by the results of the attitudes survey, follow-up survey, and interviews. The second half of the chapter provides a discussion of the broad trends revealed in the findings. These include findings on the potential of corpora to assist students with error correction, students' attitudes towards corpus use, reasons for continued use, the differences between WRI and ENG students, and students' desire for more training.

Results

Empirical Study

This section provides an account of the data collected from the empirical study. The instrument used to collect this data was the worksheets students filled in for the error correction assignment. There were 99 students who completed worksheets. Some students only filled in one, while others filled in two. In total, the WRI students produced 59 worksheets and the ENG students produced 119. However, two of the WRI worksheets were unusable as one looked at an error that had already been corrected in the student's paper, while the other was not addressing any specific error. Therefore, only 57 error correction forms from WRI were analyzed. To make an even comparison, 57 worksheets from ENG were randomly selected (19 from each section) using numbers generated by random.org, an online random number generating service.

The data from the empirical study were analyzed from two perspectives. First, students' worksheets were analyzed according to whether the student fully resolved the error s/he had set out to correct. The criteria for determining whether an error was fully resolved are described below. The second method of analyzing the data was to look specifically at the types of errors that were and were not successfully corrected. These different ways of looking at the data provide different types of information. The first provides a holistic picture of the number of students who improved their essays by resolving grammatical errors; the second provides information about the types of errors that may be more amenable to correction using a corpus.

Overall Correction of Errors

Out of the 57 worksheets analyzed from WRI, 41 contained successful error corrections and 16 contained unsuccessful corrections, resulting in a 72% rate of success among WRI students. Likewise, the number of successful corrections among the 57 ENG worksheets that were analyzed was 44, while the number of unsuccessful corrections was 13. The success rate among ENG students, therefore, was 77%. These percentages are slightly higher than those found by Gaskell and Cobb (2004), whose students made successful corrections between 60 and 70 percent of the time when consulting a corpus, and Gilmore (2008) whose students improved 61% of errors after consulting a corpus.

In order for a worksheet to be counted as a successful error correction, the student must have resolved the error(s) in the underlined portion of the sentence. For instance, in one student's original sentence, "I wanted going for the trip," the words "going" and "for" had been underlined. In her new sentence, "I wanted to go for the trip," the student successfully corrected the verb form, "going," but not the preposition, "for." Had only the word "going" been underlined, the student's correction would have been considered successful. However, since the full underlined portion of the text was not corrected, the entire correction was considered unsuccessful.

Another way students' error corrections might be counted successful was if they deleted the problematic form and replaced it with something completely different but still correct. For instance, one student had written "through out," which is a spelling error, rather than "throughout." Rather than attempt to correct the spelling error, the student replaced "through out" with "during." This error correction was considered successful.

As the results indicate, the number of successful attempts far outweighed the number of unsuccessful attempts for both levels of students. We can conclude from these numbers that when students' attention was drawn to a problematic form, they were more likely to correct it successfully than unsuccessfully. We cannot conclude that the corpus alone was responsible for the high number of successful corrections, as certain errors may have been typos, or things students could easily correct without the corpus (though precautions had been taken to reduce the likelihood of this, as discussed in Methodology).

One the other hand, several errors could have been ones the students had some idea about how to correct, but were not one-hundred percent sure about. Indeed, the fact that students made the error in the first place is a good indication that they did not know the correct form to begin with. This is why O'Sullivan and Chambers (2006) point out that "The role of the concordancer is therefore possibly to confirm a hunch" (p. 58) rather than only provide information that is new to the learner. The high number of successful corrections should not be discounted simply because some students may have only used the corpus to confirm hunches. As Yoon and Hirvela (2004) point out, students experience increased confidence when they are able to check their language hypotheses against the corpus "and see whether their linguistic choices and phrasing were correct" (p. 276).

Types of Errors

An analysis was conducted of the different types of errors which were successfully and unsuccessfully corrected in order to see whether certain error types tended to be more successfully corrected, suggesting that they might be more amenable to corpus correction. For this analysis, the results of WRI and ENG were combined, since it is the error types which are the main concern here. For a specific error type, such as spelling errors, the total number of successful corrections, unsuccessful corrections, and ignored or deleted errors (i.e., an error in which the student deleted the problematic form rather than attempting to correct it) was recorded. For instance, the number of successful corrections in the category of *spelling* reflects the number of successful spelling corrections made by WRI students plus the number of successful spelling corrections made by ENG students. The list of error types and their respective success rates is displayed in Table 2.

Although there were 57 worksheets analyzed from each level, the number of errors becomes slightly greater when we look at individual error types. This is due to the fact that one worksheet might contain more than one error. Going back to the earlier example, "I wanted going for the trip," there are two types of errors: a word form error (going), and a preposition error (for). Although the overall correction was unsuccessful (the student wrote "I wanted to go for the trip"), the correction of "going" was counted as "successful," while the correction of "for" was counted as unsuccessful.

Table 2. Types of Errors

Type	# Attempts	# Successful	# Unsuccessful or Deleted	% Successful
Word choice	32	22	10	69
Word form	25	18	7	72
Preposition	24	19	5	79
Determiner	11	8	3	73
Spelling	6	1	5	17
Extra word(s)	6	5	1	83
Missing word(s)	5	4	1	80
Transitivity	4	4	0	100
Capitalization	3	2	1	67
Word order	3	2	1	67
Punctuation	2	1	1	50
Part of speech	2	1	1	50
Total	123	87	36	70

In addition, for the analysis of error types, "unsuccessful" and "deleted" corrections were placed in the same category. The reason for combining the two was to capture the fact that if a student ignored or deleted a form, it possibly meant that the form was less amenable to correction using the corpus. For instance, the previously mentioned example of "through out," which was considered a successful error correction in terms of the overall success rate, was considered unsuccessful as far as error types, the reason being that the student did not resolve the spelling error by changing "through out" to "throughout." It must also be acknowledged, however, that the student might not have known that "through out" was a spelling error; had he known, he may have had more success using the corpus and would not have had to replace it with a different word altogether.

Successful Corrections

The three categories with the highest number of successful corrections were *word choice* with 22, *prepositions* with 19, and *word form* with 18. These three categories were also the top three attempted categories with 32, 24, and 25 attempts, respectively. The *word choice* category encompasses any error in which the student used the wrong word(s) or inappropriate-sounding words. It contains all parts of speech except prepositions and determiners, which were each put into their own categories. Therefore, while the *word choice* category encompasses nouns, verbs, adjectives, adverbs, pronouns, and conjunctions, the *prepositions* category contains only prepositions, making prepositions the single largest word class that students corrected successfully. Lastly, the *word form* category includes morphological errors in verb endings, word forms such as noun or adjective form, and singular/plural.

The large number of word choice improvements, including prepositions, demonstrates the value of using the corpus to obtain suitable collocations. Students are able to type in a "base" word and find out which words collocate with it. The success rate in the category of *word choice* was 69%, meaning that 69% of word choice errors were successfully resolved and were not ignored or deleted. The success rate in the *prepositions* category was noticeably higher, at 79%. These high success rates indicate that students were able to resolve preposition and word choice errors more often than not, without having to resort to deleting or ignoring the problematic form. Interestingly, the fourth most successfully-corrected error type was determiners, also a type of word choice error, with eight successful corrections. The success rate of determiner corrections was 73%.

The high number of successful preposition corrections is not a surprise given that other studies have found prepositions to be highly amenable to corpus correction (O'Sullivan & Chambers, 2004; Wu, Witten, and Franken, 2010, Yoon & Hirvela, 2004). One reason prepositions are often corrected successfully is because corpora usually contain a large number of examples of preposition constructions (Wu, Witten, & Franken 2010; Yoon & Hirvela, 2004). In addition, the search required for finding the appropriate preposition is a simple one since the student usually does not have to worry about intervening words in between the base word and the preposition. Examples of successfully-corrected preposition errors from this study (with the

original error in brackets) include "share a common interest in" [share a common interest of working to create a good result], "responsible for" [responsible to assign the tasks] "influential on," [being influential to others], and "devoted to" [she fully devoted herself by looking after the house].

Word form errors, which had the third-highest number of error corrections, had a success rate of 72%. With these types of errors, it is possible that students used the corpus as a hypothesis-testing tool where they already had the correct answer in mind. For instance, one student wrote "strong person interest" when it should have been "strong personal interest." The student typed "personal interest" as his search terms, suggesting that he knew the correct form from the beginning. Although the student may not have needed the corpus to correct the error, he was still able to gain confirmation from the corpus that "personal interest" was the correct form.

The remaining error categories were not highly attempted compared to the already-mentioned error types. Three of these categories, *extra word(s)*, *missing word(s)*, and *transitivity*, had relatively high success rates. The first category, *extra word* errors, is similar to word choice errors in that unnecessary word errors are a kind of collocation error where the student used an odd word combination. However, for these errors the student simply needed to delete the words that did not belong there, rather than find an alternative word. *Missing word* errors, also a type of collocation error, were those in which the text needed an additional word or phrase in order to be correct. For example, one student fixed his original sentence, "emphasize on the team qualities," by looking for collocations of "emphasize." The student added "the importance of" to the phrase, resulting in the more native-like phrase, "emphasize the importance of team qualities." Going along with missing words, *transitivity* errors were those where the student did not provide an object for the transitive verb [I enjoyed a lot]. Out of the four transitivity errors attempted, all four were corrected successfully.

The remaining error types include *word order*, *part of speech error* in which a word was used as the wrong part of speech, such as a noun used as a verb, *capitalization*, *punctuation*, and *spelling*. These categories were attempted relatively infrequently, all with moderate success with the exception of spelling in which the success rate was quite low. One of the interesting corrections made by a student in

WRI was in the *part of speech* category. The student had used "upbringing" as a verb, saying, "We all associate with upbringing her." Upon searching it in the corpus, the student was able to see that "upbringing" is used a noun. Despite not being able in his worksheet to articulate the reason behind his change, the corpus still helped the student make a successful correction with regard to the use of "upbringing" [we all associate to help with Laith's upbringing].

Unsuccessful Corrections

Corrections classified as "unsuccessful" include errors that were not successfully corrected, or that students chose to resolve by deletion. It was no surprise that the highest number of unsuccessful corrections occurred in the most highly-attempted categories: word choice, prepositions, and word form. Out of these categories, there were 10 unsuccessful corrections from the word choice category, 5 from the preposition category, and 7 from the word form category. Since certain categories were attempted more than others, it is more revealing to look at the percentage of successful corrections for each error category. From this standpoint, the spelling category was the most unsuccessfully-corrected error type, with a success rate of only 17%. That is, out of 6 attempts by students to resolve spelling errors, 5 were unsuccessful or deleted/ignored.

Based on these figures, it is reasonable to conclude that spelling errors are not amenable to the error-correction method employed in this study. Since errors were only underlined and no explanation was provided, students may have been confused about what the error was. With spelling errors, what often seemed to prevent a successful correction was the student not knowing why the text had been underlined. Three examples where students either ignored or deleted the spelling error were "through out," "life style," and "hangout" used as a verb. Had the students known the error was underlined because of spelling, they may have had better success using the corpus to correct their error.

Another problem one student ran into with spelling was that the corpus itself contained spelling errors. The student's original error was that he used "weather" where he should have used "whether." Realizing that it was a spelling error, the student typed "wether" into the corpus to see if that was the correct spelling. Unfortunately, the corpus contained examples where "wether" had been used in place

of "whether," so the student erroneously assumed that "wether" was the correct spelling.

Another example of the corpus misleading the student was with a word order error. The student had written "one member may <u>be even</u> held accountable" when the more native-like or common word order would have been "even be." The two word orders carry slightly different meanings, the former of which is only used when the writer wants to express special emphasis. Upon searching "be even" in the corpus, the student found the sentence "you have to have been dead for 25 years before it can <u>be even</u> considered," which caused him to believe that his original phrasing was acceptable. Although the example sentence the student found carries special emphasis, he did not pick up on it. The student might have been able to determine that "be even" is the less common form had he thought to check the relative frequencies between "be even" and "even be." In addition to "be even" being less common than "even be," the number of "be even" examples can be reduced even further my omitting examples of "be even + comparative adjective." Because the student did not utilize all the corpus analysis techniques available to him, and may not have even realized that his was a word order error to begin with, the corpus was not a help in this instance.

This example raises another issue that might prevent students from making a successful correction--not having "leads," or hypotheses, for how to correct an error. In the word order example, it seems that the student did not have a lead as to why the words "be even" had been underlined. Had the student known that this word order was marked and carried special emphasis, he may have been better able to address the error using the corpus. On the other hand, an example of a student who had a lead in the right direction is the student who searched "wether" in the corpus after seeing that "weather" had been marked in his paper as incorrect. His lead was that "weather" was a spelling error in the context he had used it in, and therefore he needed to search alternate spellings in the corpus.

Besides not having leads, not knowing why the text has been underlined, or the corpus containing a wrong or misleading form, another factor behind unsuccessful error corrections was that the student used a model sentence that did not match the sentence s/he was trying to correct (Kennedy & Miceli, 2001, p. 85). Taking an example from the spelling category, one student had written the music genre R&B as

"r n b." He then searched "r n b" in the corpus and got an example that read, "the school nurse in a child's education. # Eileen C. McGrath, R. N. B. S." Here, the abbreviation is referring to a nurse's title--not the music genre. The student did not realize that the example he had obtained from the corpus was not at all related to the error he was trying to correct. This student's search strategy also demonstrates a lack of leads, since the student typed in his original spelling, "r n b," rather than testing out alternative spellings.

Another example of the importance of selecting appropriate models from the corpus comes from an unsuccessful preposition correction. The student's original sentence read, "I was born at 11/2/1993." The model sentence he cited from the corpus was, "I was born in 1926." Therefore, he assumed that in order to correct the sentence he needed to change the preposition from "at" to "in." The problem with the model sentence he selected was that the date was written as the year only, rather than as day, month, year, as he had written it in his original sentence. In this case, the appropriate model would have been a sentence where the date was written as day, month, year. This mistake is important as it highlights the fact that even for preposition errors, which students are typically able to correct with the help of a corpus, the corpus is not fool-proof. It also serves as a reminder of the need to teach students to choose carefully the sentences they select as models.

In summary, the data from the empirical study was analyzed in two ways: first, each attempted error correction was analyzed according to whether it was successfully corrected or not. Second, the data was analyzed according to error types to see if certain errors tended to be more successfully corrected than others. In terms of overall error correction, the proportion of successfully corrected errors was 72% from WRI and 77% from ENG. Regarding error types, *word choice*, *prepositions*, and *word form* had the highest number of successful corrections, respectively. Other error types which had very high success rates (above 80%), but were not highly attempted, included *extra word(s)*, *missing word(s)*, and *transitivity*. The remaining error categories had low to moderate success rates and were infrequently attempted: *word order*, *part of speech*, *capitalization*, *punctuation*, and *spelling*. Of these categories, *spelling* had the lowest success rate with 17% successful corrections (one successful correction out of six attempts).

Attitude Survey

The first survey administered to students consisted of eight questions, most of which were designed to elicit students' attitudes towards corpus use. The major issues covered by the survey were whether students perceived the corpus as useful, students' confidence in their ability to use the corpus, how successful students felt they were in using the corpus, and whether they believed they would continue to use the corpus. The survey was given to students directly following their completion of the error correction assignments, which means that the survey measured the attitudes of students who had only just been introduced to the corpus less than a week before.

The first question was designed to uncover the tech savviness of the students. It asked students whether they considered their computer skills to be "average," "better than most," or "weaker than most." All of the students from WRI and 98% of the students from ENG responded with either "average" or "better than most." This indicates that nearly all, if not all, of the students were regular computer users. The second question asked students whether they had ever used a corpus before. All students except one, who was from WRI, answered that this was their first time using a corpus. One student from ENG did not answer. This indicates that nearly all of the students were new to corpus consultation and were all "beginner" corpus users.

Perceived Usefulness

To elicit their perceived usefulness of the corpus, students in WRI were asked whether they agreed with the statement, "I think the corpus is a useful tool for students wanting to improve their English." Eighty-one percent either agreed or strongly agreed with the statement. Only 19% disagreed or were unsure. This suggests that the majority of students in WRI considered the corpus a useful tool for improving their English at the time of taking the survey. Naturally, students' perceptions about whether or not the corpus is useful could change as they use it more and become more familiar with the tool.

In addition, students from WRI and ENG were asked whether they agreed with the statement, "I think the corpus is a useful tool for me." This question specifically asked whether students felt the corpus was a useful tool for them as individuals, rather than for English language learners as a whole. The data shows that the majority of students, 84% from WRI and 85% from ENG, either agreed or strongly agreed that the corpus was a useful tool for them as individuals. However, there was more

enthusiastic agreement among the WRI students, as 29% answered with "strongly agree" while only 8% of ENG students strongly agreed. These results are displayed in Table 3.

Table 3. I think the corpus is a useful tool for me (% of responses)

	SA	A	NS	D	SD
WRI 001	29	55	10	3	3
ENG 207	8	77	11	3	0

*Percentages in this table may not add up to 100 due to rounding.

Confidence

Question six of the survey asked students to rate how confident they felt in their ability to use the corpus. As a whole, students in WRI were more confident than their counterparts in ENG, as 81% of students in WRI either agreed or strongly agreed that they felt confident, while 66% of ENG students agreed or strongly agreed. Still, 66% represents a solid proportion of students who felt confident in their corpus abilities. It also seems that ENG students were more cautious about labeling themselves as confident users since 27% of ENG students, as compared to 10% from WRI, responded to the statement with "not sure." Perhaps these students were unsure how to rate their ability, or were neither confident nor unconfident about their ability. These results are displayed in Table 4.

Table 4. I feel confident in my ability to use the corpus (% of responses)*

	SA	A	NS	D	SD
WRI 001	6	74	10	3	6
ENG 207	23	44	27	6	0

*Percentages in this table may not add up to 100 due to rounding.

Training and Confidence

The third question on the survey asked whether students had attended both training sessions. This question was asked in order to see whether there was a correlation between attending the training sessions and students' level of confidence using the corpus. There was no correlation, however, as some students who attended

both training sessions were not confident in their ability, while others who had not attended both sessions were confident in their ability. For instance, 83% of the WRI students and 91% of the ENG students who did not feel confident about their ability to use the corpus had in fact attended both training sessions, while there were also some students who had not attended both training sessions but felt confident about their corpus abilities nonetheless. The fact that no correlation was found between the amount of training students attended and their confidence using the corpus suggests that other factors may have determined students' confidence, including their predisposition to using the tool, and natural ability.

Students' Success Finding What They Needed

The fourth question on the ENG survey was aimed at finding out how successful students were at finding what they were looking for in the corpus. It asked students to respond to the statement, "The corpus provides answer to my questions" with either "most of the time," "sometimes," or "rarely." The majority--58% of students--responded with "sometimes." Thirty-five percent responded with "most of the time" and 6% responded with "rarely." The fact that the majority of students selected the middle answer indicates that most students were neither overly successful nor overly unsuccessful in finding what they needed in the corpus. Still, there were more students who experienced success most of the time (35% or 22 students) than there were students who experienced success rarely (6% or 4 students). This may suggest that students could benefit from more training, that the nature of their errors and/or their language proficiency obstructed their ability to receive help from the corpus, or that the corpus itself did not contain what they needed.

Hypothesized Continued Use

Question seven asked students if they thought they would continue to use the corpus. The same percentage of students--68% from WRI and 68% from ENG--either agreed or strongly agreed. However, the proportion of students who strongly agreed from both classes (13% from WRI and 15% from ENG) was significantly smaller than the proportion who "agreed" only (55% from WRI and 53% from ENG). This suggests that most of the students who thought they would continue to use the corpus were not overly enthusiastic about it. Still, there were far more students who planned to continue using the corpus than did not plan to continue using it, as the proportion who answered "disagree" or "strongly disagree" was 10% (three students) from WRI,

and 3% (two students) from ENG. Four out of the five students who did not plan to continue using the corpus had indicated a lack of confidence in their corpus use abilities, suggesting that lack of confidence was a large factor in whether students planned to continue using the corpus.

The proportion of students who responded with "not sure" to whether they would continue to use that corpus was also substantial: 23% from WRI and 29% from ENG. There are a number of reasons students may have been unsure about whether they would continue to use the corpus. Some may have felt that their English skills were already strong, or otherwise did not have a strong motivation to improve their written English. Others might not have been convinced that the corpus itself was useful. In addition, a number of students who were unsure whether they would continue to use the corpus had indicated that they were not confident in their corpus abilities. This lack of confidence may be what was leading some students to question whether or not they would continue to use the corpus. These results are displayed in Table 5.

Table 5. I will continue to use the corpus in the future (% of responses)

	SA	A	NS	D	SD
WRI 001	13	55	23	3	6
ENG 207	15	53	29	3	0

Positive Survey Comments

The final question of the survey invited students to leave additional comments. Most of the students who left comments had positive things to say about the corpus. Some of the prototypical types of comments students left included "It is a very useful technique that allows us to learn about the most common words used interms [*sic*] of statistics" and "I think it will really help me," where the first comment was left by an ENG student and the second by a WRI student. Other students commented on specific reasons the corpus was useful, such as enabling them to look up phrases, see how a specific word is used, develop correct sentences, and discover collocations. In addition, three students (all from ENG) wrote that they found the corpus "interesting," and one from ENG wrote that it made him or her more confident. Unlike in Gilmore's (2008) study, none of the students specifically mentioned that the corpus helped them to be more autonomous or self-dependent, although this notion was implicit in many

of their comments, such as "it help [*sic*] the students to correct their errors and to check what they are writing to make sure its [*sic*] correct."

In addition, there were students who expressed a desire for the corpus to be introduced to more people. Such comments suggest that these students really found the corpus to be useful. Of the students who left such comments, one was from WRI and two were from ENG. The student from WRI suggested that corpora be introduced to all AUS students. One of the students in ENG suggested introducing corpora in English 204, a required research methods course for undergraduates. The other student from ENG simply wrote, "Very Useful [*sic*] website, just needs to be popular."

A few students also commented on how COCA was superior to other search engines, namely Google, due its specific built-in search functions. As one ENG student wrote, "It has these specific searching abilities that is mutual to my point and helps me with similar ideas." Another student pointed out that the corpus allowed him to see when various texts were written, and therefore determine whether a certain usage was "current." Finally, one student commented on the fact that COCA provides users with access to reliable sources, unlike Google which often directs users to forum posts. As the student put it,

> In the past, I have used google [*sic*] to find out if a phrase I am using is correctly used, but now I have a much better tool. Google would often refer me to forum posts, which arent [*sic*] great sources, whereas the corpus is allowing me to compare my writing with texts authored by established and accomplished writers. I will definitely be using the corpus instead of Google from now on.

This comment suggests that the student was already using previously-written texts to judge the acceptability of his/her writing, and that the corpus is a welcomed resource that this student can add to his/her arsenal of writing tools.

Negative Survey Comments

Other comments pointed out drawbacks of using the corpus. For some, the biggest drawback of COCA seemed to be the complexity of the site. One student from ENG wrote that Google was more appealing simply because it was easier to use. Likewise, another ENG student wrote, "I feel that many of the things we used corpus for in the session could have been found using Google or some other search engine,"

in other words, implying that there is no reason to use the corpus when Google can do the same job. On the other hand rather than faulting the tool for being too complex, a number of students simply expressed a need for more training, or at least more practice. Thus there were students who blamed the tool itself for not being to their liking, and others who believed they were the ones who needed to adapt to the tool. There were students from both camps in WRI and ENG, suggesting that a certain attitude was not specific to the level of the students.

Finally, there were a number of comments stating that although the corpus was a useful tool, students probably would not take the time to use it. As one ENG student put it, "if a research paper is about 8 pages long, writers will find it troublesome to continue resorting to the corpus for every little detail that they come across." Here, the comment seems to be more about the student's attitude towards proofreading rather than about the corpus itself, since the student seems to be opposed to checking every detail. This was in line with other comments in which students mentioned that they would only be inclined to use the corpus for really important assignments, suggesting that they are not very concerned with the accuracy of their language for regular assignments.

Ironically, a few students also commented that going to the Writing Center and peer review are more efficient ways to proofread one's work. While that may be true, it seems these students did not understand that the point of using a corpus is to eliminate the need to seek out another person's help. Again, these comments came from both WRI students and ENG students, so no one attitude was particular to a certain level.

In summary, the attitude survey suggests that the majority of students in the study (over 80%) perceived the corpus as a useful tool and that 68% of students from both WRI and ENG believed they would continue to use it. It also found that more students in WRI (81%) felt confident about their ability to use the corpus than in ENG (66%), although these are both solid proportions of the total body of students. In addition, it was found that most ENG students (58%) found answers to their questions "sometimes" rather than "most of the time," suggesting that for these students, the corpus might not be able to provide all the answers students are looking for. Positive comments from the students included wanting to see the corpus become "popular," while negative comments included that it was easier to go to the Writing Center than

use the corpus, and that the corpus was more complex than Google, which made Google their preferred source.

Interviews

The interviews provided an opportunity for students to express what they liked, disliked, and would change about the corpus, as well as expand on what they had indicated in the attitude survey. Many students offered suggestions for features they would like to see added to the corpus. For instance, one WRI student mentioned that he would like to see a spell check feature, while one ENG student said that he would like to be able to enter whole sentences and have the corpus tell him whether or not that sentence was properly-formed. Both of these comments suggest that these two students may have believed that the corpus was a tool specifically designed for language learners. In addition, the comment about wanting the corpus to be able to analyze a sentence suggests that the student misunderstood the premise of the corpus, which is simply to provide language data for the user to interpret.

Perceived Advantages

In line with the survey comments, students who were interviewed had a number of positive things to say about the corpus. One of the things the students all agreed upon was the usefulness of being able to see multiple lines of text containing the target form. As one WRI student put it, "you have more, you will know how to use it." Another student from ENG mentioned that she appreciated having examples from different sources, as she used the source to help her decide whether the sentence could be of use to her. This finding is in accord with other studies (e.g., Chambers, 2005; Yoon & Hirvela, 2004) which have found that students appreciate having multiple lines of examples in the concordance as it helps them verify their answers. In addition, a different ENG student pointed out that the multiple lines in the corpus not only helped her correct her errors, but also gave her ideas. Although using the corpus to get ideas was not discussed in the training sessions, this is a recognized, valid use of a corpus (Kennedy & Miceli, 2001, 2010).

Another aspect students appreciated about the corpus was its "trustworthiness" in terms of containing credible sources with appropriate and accepted language usage. Students are correct to feel this way and had been told in the training sessions that this was an advantage of using the corpus. However, the students' high confidence in the corpus was concerning in that some seemed to feel that anything they found in the

corpus was appropriate to use, regardless of context. For instance, one student from ENG made the comment that the corpus was useful because "you can go with it without thinking about it...whatever shows up you can use it," and two of the students interviewed from WRI also seemed to view the corpus as a panacea. This attitude that anything that shows up in the corpus is usable may explain why so many students gave positive endorsements of the corpus in the survey. Had students been more aware of the need to be critical of what one finds in the corpus, they may not have been as likely to view the corpus as positively as they did.

Perceived Drawbacks

Although some of the students interviewed had nothing bad to say about the corpus, others seemed to feel that the corpus was a bit difficult to use. For instance, one of the ENG students mentioned that he usually did not find what he was looking for on the first try, and that it was necessary to try different combinations of search terms. He added that he did not really have time to delve in to learning about the ins and outs of the corpus, implying that he believed it was a complex tool. This student also did not really feel that it was necessary for him to use the corpus, since "Till now, Microsoft Word does the job for us." A different student, also from ENG, mentioned that the corpus would not be his "first resort" when revising a paper, and that he preferred peer review or leaving his writing for a day and returning to it later. Again, this student may have been put off by the complexity of the corpus, deciding to stick to what was familiar to him and what worked for him in the past.

Desire For More Training

When it came to training, many of the students interviewed felt that the training they had received was adequate and that it was enough just to be introduced to the corpus and taught the basic search techniques. These students stressed that the corpus was not difficult to use, especially since many of them were engineering majors and had to deal with much more complicated software. However, somewhat contradictorily, these same students also felt that they needed more hands-on time using the corpus. One student implied the she was her own best teacher, saying, "everything has its hiddens [*sic*] inside...if I had more practice, I might find other ways...something that will help me more." Everyone interviewed agreed that they would become better corpus users with practice.

There were, however, a few students who felt that additional training would be helpful. One student pointed out that the training had not given her and her classmates much time to digest what they had been taught. She suggested extending the training period across two weeks, and also having one session devoted to letting students ask questions. A different student replied that she was open to receiving more training, as perhaps she would learn something she would not have been able to figure out on her own. Lastly, there was one student who emphasized wanting to learn about applications of the corpus. Perhaps what he meant by "applications" was wanting to see the corpus in action; in other words, to see demonstrations of how the corpus could help resolve errors in actual students' writing.

Grammatical Knowledge

One of the important considerations the interviews revealed was whether grammatical knowledge plays a role in a student's ability to use the corpus. Various students who were interviewed mentioned how the corpus helped them find the appropriate preposition or verb. The fact that students mentioned these terms suggests that they were drawing on their knowledge of grammar to some extent when using the corpus. Further research would help to shed light on the extent that grammatical knowledge facilitates successful corpus use. It is also possible that students are not always consciously aware of the grammatical rules governing a construction, yet are still able to benefit from a corpus.

Resource Usage

When introducing students to a resource such as a corpus, it is interesting to know whether the students themselves were already users of resources such as dictionaries and other reference tools. As mentioned already, several students were accustomed to taking their papers to the Writing Center; they also used peer review or giving one's paper to a friend or family member to read. Students also mentioned that they used computer resources, which consisted of the editing tools in Microsoft Word such as spell check, grammar check, and the built-in thesaurus, as well as various internet resources including online dictionaries and thesauruses. Other students said they utilized search engines like Google by typing a word or phrase into the search bar which would take them to online dictionaries or forums. Others also mentioned using Google Translate. Only one of the students who was interviewed admitted to not using any online resources. When asked why, he said it was because he does not

like computers, although he said that despite not liking computers, he did enjoy using the corpus. Interestingly, none of the students said they used paper or book resources, which suggests that today's students are much more likely to consult electronic resources than books.

Overall, a number of interesting findings came out of the interviews. One was that, based on their suggestions for ways to improve the corpus, some students may have believed that it was designed specifically for language learners. Another finding in accord with the literature (e.g., Chambers, 2005; Yoon & Hirvela, 2004) is that students appreciated having multiple lines of examples in the concordance as it helped them verify their answers. None of the students found the number of results "overwhelming," although students in other studies (e.g., Liu & Jiang 2009) have mentioned this as a drawback of corpora. The interviews also suggested that a few students saw the corpus as a panacea, and suggests that more should be done in the training to teach students to use discretion when selecting examples. Lastly, all of the students interviewed felt that the training had enabled them to use the corpus, but a few would have appreciated more training. However, there were other students who were content to learn more about the tool on their own and did not feel that they needed more training.

Follow-Up Survey

The follow-up survey served two purposes: to see whether students desired more training, and to find out how many students had taken up use of the corpus. The survey was administered six weeks after WRI students had been introduced to the corpus, and five weeks after ENG students had. Some ENG students took the survey a week after it opened, making it six weeks since they had been introduced to the corpus. In total, 77 students took the follow up survey: 31 students from WRI and 46 students from ENG.

Desire For More Training

One of the questions in the survey asked students whether or not they would like to receive more training in how to use a corpus. Of the WRI students, 58% agreed or strongly agreed, 29% were not sure, and 13% disagreed or strongly disagreed. From the ENG students, a smaller proportion than the WRI students (43%) agreed or strongly agreed, a larger proportion than the WRI students were not sure (43%), while approximately the same proportion (14%) disagreed or strongly disagreed. These

results indicate that a greater proportion of WRI students wanted to receive more training, while a large proportion of the ENG students were undecided about whether they would like to receive more training or not. These results are displayed in Table 6.

Table 6. I would like to receive more training in how to use a corpus (% of responses)

	SA	A	NS	D	SD
WRI 001	3	55	29	3	10
ENG 207	4	39	43	7	7

Uptake

Students were also asked whether they had used the corpus since being introduced to it six weeks before. The proportion of students from WRI who claimed to have used the corpus independently was 52%, while the proportion of students from ENG who claimed to have used it was 41%. These figures suggest that a slightly greater proportion of WRI students took up use of the corpus than ENG students. The proportion of uptake from both levels, however, is quite sizable given that continuing to use the corpus was completely optional.

In addition, the students who had taken up use of the corpus were asked whether they had used it "five times or less" or "more than five times." Again, the WRI students demonstrated greater use of the corpus, with 44% answering that they had used the corpus "more than five times" since being introduced to it. In contrast, only 5% of ENG students indicated that they had used the corpus more than five times.

Reasons Behind Uptake

Students who indicated that they had continued to use the corpus were asked what factors led them to continue using the corpus. Their responses indicate, not surprisingly, that students who continued to use the corpus did so because they perceived it as being helpful to them. Specific reasons WRI students gave for using the corpus include that it was helpful and easy to use. Other comments from WRI students included "To make sure of using the right words" and "showing words in context." Surprisingly, several WRI students also indicated that they used the corpus as a vocabulary tool, even though the corpus had been presented to them as a way of looking up collocations, not definitions. Their responses included comments like "looking for meaning of unfamiliar words," "learning new words," "helping with my

vocabulary," and "finding out the meanings of strange words." Perhaps these comments are revealing of students' metalanguage and the way they think about writing. Maybe to certain students, "learning new words" is synonymous with "seeing how a word is used." Alternatively, it is possible that students were in fact using the corpus in place of a dictionary.

Reasons behind ENG students' continued use of the corpus also included comments about how the corpus was "helpful" and "useful." Others also mentioned its ease of use, leaving comments such as "easy and simple," "easy to use," and "a practical way to ensure the kind of language I use in my daily basis assignments." Other students named specific ways they had used the corpus, which included looking up prepositions, finding synonyms of words, and checking technical language. Students also gave more general comments such as "to make sure about my language usage" and "to use appropriate words," and "interesting to see whether or not professional writers use the same phrasing I do." The common thread running through all these comments seems to be that students viewed the corpus as an authoritative resource that could tell them whether or not their language usage was appropriate.

Reasons Behind Non-Uptake

The students who had not taken up use of the corpus were asked to comment on the factors that deterred them from using the corpus. Because this was a required question, a number of negative attitudes came out which had not come out on the initial survey where leaving a comment was optional. For instance, two of the comments from WRI students said that they did not find the corpus useful. In addition, comments from ENG students included "I did not find it to be very usefull [*sic*]," "I think it is a waste of time," and "I found it to be really time consuming and unneccassery [*sic*]." Another person commented that s/he did not think s/he "would get much useful result from using it" since this person felt it was only minimally helpful compared to the time needed to search the corpus. Another attitude among some students was that other sources were more useful than the corpus. For instance, one student commented that he preferred using a dictionary.

However, many of the reasons students cited for non-uptake had to do with not having a need for the corpus. For instance, among the WRI comments, some students simply wrote that they did not have any difficult assignments in which they felt that they needed to use the corpus. Several students from ENG also cited a lack of need,

45

although it was unclear whether this was due to not having any difficult assignments, or due to their perception that their English skills were already good enough.

Another reason students cited for non-uptake was that they found the corpus to be too complicated or did not have a clear understanding of how to use it. The most common complaint from students, however, was that using the corpus was time-consuming. For some students, the time-consuming nature of the corpus was due to its complexity and/or to their lack of understanding in how to use it. For others, the corpus was time-consuming to use because it did not offer very much payoff, or in one student's words, "it takes much time for one to figure out tiny bits."

In summary, a significant proportion of students (58% from WRI and 43% from ENG) desired more corpus training. This suggests that students were not as confident in using the tool as they would like to be; it also suggests that they believed it was a tool worth knowing how to use. The uptake from students was also fairly high, with 52% of WRI students and 41% of ENG students claiming that they had used the corpus since first being introduced to it. Students who had taken up use of the corpus listed that they had done so because they found the corpus useful. Conversely, those who had not taken up use of the corpus listed reasons such as not having an occasion in which they needed to use it, not believing it was useful, or finding it too complicated and time consuming. Other students expressed a desire to use the corpus, but indicated that they needed some time to figure it out, which they had not had.

Discussion

Empirical Study

Two overall conclusions can be drawn from the empirical study. First, given the high error correction success rates (72% from WRI and 77% from ENG), it can be concluded that when students' attention was drawn to a problematic form, they were more likely to correct it successfully than unsuccessfully. Second, there was not a very big difference in the success rates of WRI and ENG students. This indicates that WRI students, despite being at a lower level than the ENG students, had nearly the same success rate as their higher-level counterparts when altering problematic forms in their writing.

However, we cannot conclude that the corpus alone was responsible for the high number of successful corrections, as certain errors may have been things students

could have corrected without the corpus. By the same token, it is possible that even if students knew the correct form before consulting the corpus, the corpus helped them confirm their hunch and make them more confident in their answer. Finally, it is also possible of course that the corpus provided students with new information that helped them correct their error when they otherwise would not have been able to. In order to know whether students knew the answer before consulting the corpus, one could observe individual students as they were using the corpus and have them verbalize their thought processes.

Of course, there were also instances when students were unsuccessful in correcting their work, which means there were times when the corpus was not able to provide enough support to help students correct their errors. One reason the corpus may not have been able to help certain students was because students did not understand the error and therefore could not formulate an effective search strategy for resolving it. The students may not have known why the error was underlined to begin with, or else may not have known of any ways to resolve the error (lack of leads). Another reason the corpus may not have been able to help students was because the corpus itself contained errors or odd forms, such as "wether." Lastly, students may have misused the corpus by selecting inappropriate model sentences on which to base their corrections, as the student did who took "R. N. B. S.," a nurse's title, out of context and assumed it was the correct spelling for "R&B."

Lastly, the empirical study helped to shed light on the types of errors which may be most amenable to corpus correction. First, the error types students selected most often to correct using the corpus were word choice errors (where the parts of speech included nouns, verbs, adjectives, adverbs, pronouns, and conjunctions), preposition errors, and word form errors. These error types also had a high rate of being successfully corrected (69%, 79%, and 72%, respectively). With regard to the word choice and preposition errors in particular, the high rate of successful corrections points to the value of using a corpus to obtain suitable collocations. However, as the percentages indicate, there is still a possibility for error. One example concerning prepositions was the student who used an inappropriate model sentence and ended up saying "I was born in 11/2/1993."

Attitudes Towards Corpus Use

The attitudes WRI and ENG students displayed in the initial attitudes survey, interviews, and follow-up survey were generally similar. When asked whether they thought the corpus was a useful tool for them, 84% of students from WRI and 85% of students from ENG either agreed or strongly agreed. The proportion of students who did not think the corpus was useful was relatively small (6% from WRI and 3% from ENG) compared to the proportion who felt it was useful. In general, students appreciated the fact that the corpus provided examples of words used in context, as well as the fact that it featured trustworthy and reliable sources. A number of students also felt that it was a useful tool which needed to be introduced to all students at the university. Students also mentioned the fact that the corpus could give them ideas, as well as tell them if certain words or phrases were current.

Regardless of whether students felt the corpus was useful or not, many felt that it was difficult to use in the sense that it was a complex and rich tool. From the comments provided in the surveys and interviews, students who felt the corpus was difficult to use fell into two groups: those who blamed the tool itself for being too complex, and those who believed they were the ones who needed to adapt to the tool and become a better user. Those who fell into the former group felt that better resources existed, such as Google, Microsoft Word, online dictionaries, and going to the Writing Center. Those who mentioned the Writing Center as a more efficient way to have one's work revised clearly did not understand that the point of using a corpus is to eliminate the need to seek out another person's help.

Many also felt that using the corpus was time-consuming. Some believed that this was due to their lack of experience, while others blamed the tool itself. In addition, a recurring attitude among students from both levels was a lack of motivation to use the corpus. Several students indicated, for instance, that they would only use the corpus for really important assignments. Others felt that using the corpus was tedious, since it could only help with "minor" mistakes such as prepositions--the implication being that these errors were not worth taking the time to correct. More than one student also had the misconception that if they were to use the corpus, they would be obligated to use it to correct every mistake in their paper. For this reason, they discounted the corpus as being too tedious and time-consuming.

Continued Use

Given the high proportion of students who believed the corpus was a useful tool, one would expect a large proportion to continue using the corpus to improve their writing. Indeed, the proportion of students who continued to use the corpus even after it was no longer required was 52% from WRI and 41% from ENG. These percentages, which reflect students' actual usage, are less than students' hypothesized continued use in which 68% of students from WRI and ENG believed they would continue to use the corpus. Nevertheless, the percentages of actual uptake are quite sizable given that continuing to use the corpus was completely optional. These actual uptake percentages are also consistent with students' hypothesized continued use in various studies, in which 40-55% of students believe they will continue to use the corpus in the future (Gaskell & Cobb, 2004; O'Sullivan & Chambers, 2006; Liu & Jiang, 2009).

Regarding the discrepancy between levels, it is difficult to know why the percentage of students who took up using the corpus was greater for WRI than ENG. Perhaps it is because there was a greater proportion of students in WRI who were insecure about their language. It could also have been that due to their low proficiency level, some WRI students may have misread the question as "have you ever used the corpus?" rather than "have you taken up using the corpus on your own?"

Differences Between WRI and ENG Students

Regarding WRI and ENG students' ability to benefit from the corpus, it was found that both groups benefitted to a similar degree. That is, WRI students only made slightly fewer successful corrections (41) than ENG students (44). These results are surprising considering that one would expect advanced students to be more proficient in the language and have greater prior knowledge--i.e., understanding of the error to be resolved and an idea of possible alternatives. One explanation which may account for the similarity among the two groups is that students in WRI likely chose errors which were easy for them to correct, and which they had prior knowledge about. The fact that the corpus possibly did not help WRI students with more "advanced" errors is not necessarily a bad thing. In fact, this explanation corroborates what previous researchers (i.e., Breyer, 2009; Gilmore, 2008) have said about how corpora help students learn what they are ready to learn.

Although WRI and ENG students generally exhibited similar attitudes, students from ENG were more cautious about strongly agreeing that the corpus was a useful tool. One possible reason ENG students showed less enthusiasm for the corpus was that they probably consider themselves more proficient in English than their counterparts in WRI. There was also a noticeable difference in the percentage of ENG students who felt confident in their ability to use the corpus (66%) versus the percentage of WRI students who felt confident in their ability to use the corpus (81%). The ENG students may have been less confident in their ability due to the fact that they were introduced to more advanced features (namely, the collocation feature), while WRI students were not. In addition, ENG students may have been more critical of their abilities, causing them to rate themselves as being less confident.

Indeed, the high percentage of WRI students who felt confident in their ability to use the corpus may be a cause for concern given that there were more students who felt confident in their ability to use the corpus than made successful error corrections. Students may have felt overly confident, believing that anything was correct as long as it was in the corpus. One ENG student's comment, though already mentioned, is worth repeating as other students may have felt the same way: "you can go with [any example from the corpus] without thinking about it...whatever shows up you can use it." Thus, the increased confidence students reported in this study and others such as Yoon's (2008), may actually reflect a false sense of security created by not being critical enough of what one finds in the corpus.

Training

There was not a strong correlation between attending both training sessions and the level of confidence students possessed when using the corpus. It seems that students' predisposition to corpus consultation and natural ability with computers may have been more of a factor in determining their confidence level than whether or not they attended both days of training. In addition, the difference between attending one day of training versus two days of training is not as large as the difference between attending one day of training and, for example, one week of training. In other words, the amount of training students received was probably too small to significantly alter their natural ability to use the corpus and catch on to new software.

In addition, there was quite a large proportion of students from both levels who would have liked to have had more training (58% from WRI and 43% from ENG).

This suggests that for a large proportion of students, two class sessions of training was insufficient. In addition, the comments on the follow-up survey suggest that many students did not take up use of the corpus due to its complexity. Therefore, offering more training might be one way to increase the numbers of students who take up use of the corpus. Given one student's suggestion of stretching the training out across two weeks, it might work well to offer ongoing training that is integrated into the course, as done in previous studies (Kennedy & Miceli, 2001; Yoon & Hirvela, 2004; Yoon, 2008).

To conclude, this chapter has presented the findings of the research and a discussion of these findings. The results of the empirical study suggest that corpus consultation can help learners correct errors in their writing. The attitude survey found that the majority of students believed the corpus was useful, while the interviews shed more light on why students felt it was useful. In addition, the follow-up survey suggested that many students had taken up use of the corpus. Finally, feedback from the students suggests that around half of them would appreciate the opportunity to receive more corpus training.

CHAPTER FIVE: CONCLUSION

This chapter provides an overview of the findings from the study and answers the research questions. In addition, implications for teachers wishing to implement corpus consultation are discussed. Finally, limitations of the study and suggestions for future research are provided.

Overview of Findings

The findings from this study suggest that students can benefit from direct corpus consultation and that a high proportion of students found the corpus useful. This confirms findings from previous studies (e.g., Watson-Todd, 2001; O'Sullivan & Chambers, 2006; Wu, Witten, & Franken, 2010; Gilmore, 2008; Gaskell & Cobb, 2004; Yoon, 2008) that corpus consultation can help students improve their writing. No significant differences were found between WRI and ENG students in terms of their ability to use the corpus and their attitudes towards it. Therefore, it is recommended that tertiary level ESL writing teachers, from the most basic levels to the most advanced, consider adding a corpus component to their classes.

Results from the study are also consistent with specific findings from previous studies, such as the utility of the corpus for helping students correct preposition errors (O'Sullivan & Chambers, 2006; Wu, Witten, & Franken, 2010). Similarly, as in Yoon's (2008) study, students in this study reported that the corpus made them feel more confident about the accuracy of their writing. This study also adds to the literature by not only asking students whether they *think* they will continue to use the corpus, but whether they actually do. It found that the actual continued usage figures (between 40-55%) were consistent with students' projected usage figures in previous studies (e.g., O'Sullivan & Chambers, 2006; Liu & Jiang, 2009; Gaskell & Cobb, 2004).

Answers to Research Questions

In answer to the first research question, *How successfully do students use a corpus for error correction after two class sessions of training?*, it was found that students were highly successful. In the majority of cases, 72% from WRI and 77% from ENG, students were able to correct problematic forms successfully. However, the role the corpus played in these corrections is unknown. While in some cases the corpus may have provided students with answers they would not have known

52

otherwise, it is also possible that students used the corpus to confirm their hypotheses about how to correct the problematic form.

The second research question asked, *What are students' attitudes towards corpus use after two class sessions of training?* Overall, students' attitudes towards corpus use were positive, with 84% of students from WRI and 85% from ENG believing that the corpus was a useful tool. In addition, students provided a number of comments in both the survey and interviews about how the corpus was a good and useful tool. In fact, some students found the corpus to be so useful that they recommended it be introduced to all students at the university. At the same time, there were also students who did not find the corpus useful. These students regarded the corpus as unnecessary or a waste of time. However, these students were in the small minority.

The third research question asked, *What are the differences between basic and advanced students' proficiency with and attitudes towards corpora?* With regard to proficiency, no significant differences were found between WRI and ENG students' ability to benefit from the corpus, as both groups made nearly the same number of successful corrections. One explanation for why WRI students were nearly as successful as their ENG counterparts is that students from WRI selected errors that were easy for them to correct.

With regard to attitudes, it was found that although WRI and ENG students' attitudes were, for the most part, similar, there were some noticeable differences. For instance, although nearly the same proportion of WRI and ENG students believed the corpus was useful, a noticeably larger portion of WRI students strongly agreed than ENG students. As discussed previously, this may be due to the fact that WRI students were more convinced than ENG students that they needed the corpus. Another interesting difference was that WRI students felt more confident about their ability to use the corpus than ENG students. This may have been due to the fact that ENG students had been introduced to more complex features than the WRI students, or it may have been because ENG students were more critical about their abilities.

The answer to the fourth question, *Do students continue to use the corpus even after it is no longer required by the instructor?*, is yes. To be more specific, 52% of students from WRI and 41% from ENG reported that they had continued to use the corpus since being introduced to it. These percentages are consistent with other

studies which found that the percentage of students who believed they would continue to use the corpus was between 40-55% (Gaskell & Cobb, 2004; O'Sullivan & Chambers, 2006; Liu & Jiang, 2009).

The final research question asked, *What types of errors, if any, does corpus consultation help students correct?* The findings from this study suggest that corpus consultation can help students with many types of errors since, in all twelve error categories in this study, there was at least one successful correction. However, from the perspective of which errors tended to be most often corrected successfully, *transitivity* errors are at the top with 100% successful corrections. In second place are *extra word(s)* errors with 83% successful corrections, and in third place are *missing word(s)* errors with 80% successful corrections. However, these categories were not highly attempted, with only four, six, and five attempts, respectively. Therefore it is difficult to generalize whether these error types would consistently have such high success rates.

In contrast, the fourth most successfully corrected error type, prepositions, was highly attempted with twenty-four attempts in total. Seventy-nine percent of these attempts were successfully corrected--that is, nineteen out of twenty-four. These results strongly suggest that corpora can assist with the correction of preposition errors. The study also found that the corpus can be useful for word choice errors involving other word classes; however, the single most successfully corrected word class was prepositions. Word form errors were another category with a fairly high success rate (72%) that was highly attempted (twenty-five times). Types of word form errors included subject-verb agreement and morpheme usage, such as whether to use -*ing* or -*tion* at the end of a noun. These errors are often lexicogrammatical in nature, since, for example, "communicating" and "communication" are both acceptable words depending on the context in which they are used.

The study also found spelling errors to be the least amenable to corpus correction, at least for these students, with only one out of six attempted spelling errors successfully corrected. One possible explanation for the low number of successful spelling corrections is that students did not know the error was a spelling error. Another possibility is that the student knew the error was a spelling error, but did not know of any alternative spellings. For instance, the student who wrote "through out"

instead of "throughout" may not have thought to check whether it should be written as all one word.

Implications for Teachers

Regarding the sufficiency of the training, the results indicate that a large proportion of students would appreciate more training, even though the two day's worth of training offered was enough to enable students to make many successful corrections to their papers. In order to appeal to students' desire for more training, however, it is recommended that teachers intersperse corpus activities throughout the semester. Another reason to consider offering more training is that students in the study who held negative views about concordancing generally felt that the corpus was overly complex and time-consuming. With more training, it may be possible to change some of these negative attitudes into positive ones. At the same time, concordancing may never appeal to certain students due to factors such as their learning style or lack of motivation.

However, precautions should be taken to ensure that students' do not overestimate the power of the corpus, as this may offer them a false sense of security. Students must be taught to critically evaluate what they find in the corpus before assuming that it is appropriate to use in their own writing. It is also recommended that teachers differentiate between plagiarism and using examples from the corpus as models. This consideration was triggered by one student's comment that "copying native speaker's words is really good for us as non-native speakers." It is important, therefore, to emphasize that the corpus is meant to illustrate patterns for students to learn from--not provide fodder for them to copy from.

Teachers who decide to integrate corpus consultation into their classes can gain useful information from this study. For instance, it is reasonable for teachers to expect over 70% of the students to have positive attitudes towards concordancing. Teachers can also better anticipate what types of errors the corpus can help their students with-- for instance, it will more likely help students with word choice and adjusting word forms than with spelling. Also, although this study implemented corpus consultation after the teacher had given feedback, teachers can also present it to students as a tool to use while they are writing. Indeed, introducing students to the corpus is more than giving them access to a new resource--it is also a way of training them to become

more autonomous learners and to adopt a different attitude towards taking control of their own learning.

Finally, there are a number of directions an institution may take in order to implement corpus use in ESL writing classes. If interest is high among teachers to learn to use a corpus, this could provide an initiative at the university to offer corpus training for teachers. If interest is low, this may suggest that corpus instruction may be better implemented through outlets such as the Writing Center or bringing in a knowledgeable corpus user into classes to train the students. For students who wish to receive more training, the Writing Center could hire tutors who are knowledgeable corpus users and also offer workshops for interested students to attend.

Limitations of the Study

One of the main limitations of the study is the fact that there were no control groups to compare the number of successful corrections made with the help of a corpus and without. Ideally, there should be one group of students who use the corpus to correct their papers, and one group who use traditional reference tools such as dictionaries, and then the number of successful corrections from both groups should be compared. Another limitation of the study concerns the timing of when the surveys and interviews were conducted. Because the first survey was administered to students days after they were first introduced to the corpus, their attitudes towards the tool may have been overly positive, or negative, due to not having a chance to become fully acquainted with the tool.

Recommendations for Future Research

Given the limitations of the present study, one of the suggestions for future research is to have a control group as part of the empirical study to test the true benefit the corpus provides over traditional or no resources. More research into the training practices of teachers and researchers would also be extremely valuable. The more ideas that become part of our collective knowledge and are tested, the better we will be able to provide effective training for our students so they may become confident and proficient corpus users.

REFERENCES

Aston, G. (2001). Learning with corpora: An overview. In G. Aston (Ed.), *Learning with corpora* (pp. 7-45). Bologna: CLUEB.

Aston, G. (1995). Corpora in language pedagogy: Matching theory and practice. *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 257-270). Oxford: Oxford University Press.

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In J. M. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15-36). Amsterdam: John Benjamins.

Bernardini, S. (2000a). *Competence, capacity, corpora.* Bologna: CLUEB.

Bernardini, S. (2000b). Systematising serendipity: Proposals for concordancing large corpora with language learners. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective: Papers from the Third International Conference on Teaching and Language Corpora* (pp. 225-234). Frankfurt am Main: Peter Lang.

Boulton, A. (2009). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL, 21*(1), 37-54. doi: 10.1017/S0958344009000068

Breyer, Y. (2009). Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning*, *22*(2), 153-172. doi:10.1080/09588220902778328

Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning and Technology*, *9*(2), 111-125. Retrieved from http://llt.msu.edu/vol9num2/chambers/

Chambers, A., & O'Sullivan, I. (2004). Corpus consultation and advanced learners' writing skills in French. *ReCALL*, *16*(1), 158-172. doi: 10.1017/S0958344004001211

Cheng, W., Warren, M. & Xun-feng, X. (2003). The language learner as language researcher: Putting corpus linguistics on the timetable. *System*, *31*, 173-186. doi:10.1016/S0346-251X(03)00019-8

Davies, M. (2008-). *The Corpus of Contemporary American English (COCA): 410+ million words, 1990-present.* Retrieved from http://www.americancorpus.org

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, *32*, 301-319. doi: 10.1016/j.system.2004.04.001

Gavioli, L. (2001). The learner as researcher: Introducing corpus concordancing in the classroom. In G. Aston (Ed.), *Learning with corpora* (pp. 108-137). Bologna: CLUEB.

Gavioli, L., & Aston, G. (2001). Enriching reality: Language corpora in language pedagogy. *ELT Journal*, *55*(3), 238-246.

Gilmore, A. (2008). Using online corpora to develop students' writing skills. *ELT Journal*, *63*(4), 363-372. doi: 10.1093/elt/ccn056

Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 359-369). London: Routledge.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Johns, T. (1986). Micro-concord: A language learner's research tool. *System*, *14*(2), 151-162.

Johns, T. (1991). Should you be persuaded - Two samples of data-driven learning materials. In T. Johns & P. King (Eds.), *Classroom Concordancing* [ELR Journal, 4] (pp. 1-16).

Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of Data-driven Learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 293-313). Cambridge: Cambridge University Press.

Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning and Technology*, *14*(1), 28-44. Retrieved from http://llt.msu.edu/ vol14num1/kennedymiceli.pdf

Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning and Technology*, *5*(3), 77-90. http://llt.msu.edu/vol5num3/kennedymiceli/

Liu, D., & Jiang, P. (2009). Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal, 93*(1), 61-78. doi: 10.1111/j.1540-4781.2009.00828.x

Meyer, C. F. (2002). *English corpus linguistics: An introduction.* Cambridge: Cambridge University Press.

O'Sullivan, I., & Chambers, A. (2006). Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing, 15*, 49-68. doi:10.1016/j.jslw.2006.01.002

Sripicharn, P. (2010). How can we prepare learners for using language corpora? In A. O'Keeffe, & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 371-383). London: Routledge.

Sun, Y-C. (2003). Learning process, strategies and web-based concordancers: A case study. *British Journal of Educational Technology, 34*(5), 601-613.

Tsui, A. B. M. (2004). What teachers have always wanted to know--and how corpora can help. In J. M. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 39-61). Amsterdam: John Benjamins.

Turnbull, J., & Burston, J. (1998). Towards independent concordance work for students: Lessons from a case study. *On-CALL, 12*(2), 10-21.

Watson-Todd, R. (2001). Induction from self-selected concordances and self-correction. *System, 29*, 91-102.

Wu, S., Witten, I. H., & Franken, M. (2009). Utilizing lexical data from a Web-derived corpus to expand productive collocation knowledge. *ReCALL, 22*(01), 83-102. doi: 10.1017/S0958344009990218

Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology, 12*(2), 31-48. Retrieved from http://llt.msu.edu/vol12num2/yoon/

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing, 13*, 257-283. doi:10.1016/ j.jslw.2004.06.002

APPENDIX A

Training Worksheet Given to Writing 001 Students

Name: _____

Date: _____

Corpus Activity

<u>Directions:</u> Use the corpus to help you answer the following questions.

1. What does "COCA" stand for? _____

2. COCA contains language from which country? _____

3. Pick a word or group of words of your choosing and answer the following questions:

    a. How many times does your word or words appear in the corpus?

    b. What is the most common word preceding your search word(s)?

    c. What is the most common word following your search word(s)?

    d. How are the results organized in the List display?
        A. By text type (magazine, fiction, newspaper, etc.)
        B. By year
        C. By year and text type

    e. Practice jumbling the results in the list display. (You can do this by going to "Sample" at the top of the page and then selecting the number of words you want in the sample (100, 200, etc.)

    h. How many times does your word appear in the following texts of the corpus?
        Fiction: _____
        Newspaper: _____
        All except Spoken: _____

4. Try to correct the following sentences using the suggested search terms. (Make sure you search in all text types except Spoken):

    a. Apart from enjoying <u>with</u> my cousins, I also spent some quality time with my grandparents. (enjoying *)

    b. Such a solution needs hard <u>working</u> and respect. (hard *)

    c. Getting good grades <u>needs</u> a lot of effort. ([v*] effort)

    d. Feedback reduces the chances of <u>mislead</u> conclusions. (* conclusions)

    e. Staying many hours in the library has resulted <u>to</u> more cohesion between group   members. (resulted *)

f. There are many fresh graduates wanting to enter the <u>work-field</u>. (work*)

g. The competitions and shows were life <u>teaching</u>. (life *ing)

h. Search for sentences that start with "because" and read through them to find out what they all have in common. (. because)

5. Which one do people usually say:
    A. at the CSU
    B. at CSU
*CSU stands for Colorado State University

6. Which one do people say more:
    A. I have met many friends
    B. I have made many friends

7. Which one is correct:
    A. Victoria English school
    B. Victoria English School
    C. victoria English School
*Note: if it's not in the corpus, search for it on the internet!

8. Look up anything in the corpus that you're curious about. Write down any interesting findings in the space below.

_____

_____

_____

_____

_____

_____

_____

_____

_____

APPENDIX B

Corpus Guide Given to English 207 Students

Corpus Search Techniques

<u>Basic Search Techniques:</u>

To do a basic search, just type a word or words into the WORDS box and press Search

Searches with punctuation: remember to treat punctuation like its own word by putting a space between it and the other words.

Wildcard search (e.g., resulted *)

Wildcard with specified part of speech (e.g., resulted [i*])

Wildcard attached to the word (e.g., work*) - good for discovering prefixes, suffixes, and compound words

Lemma search (i.e. using brackets) to find the different forms of a word, e.g., [sing]

Specifying the part of speech of a word: cause.[v*]

Filtering by text type (for example you may search all text types except Spoken)

To jumble the search results, go to Sample and select 100.

To see the expanded context of a line in the concordance, simply click on it under the "Click for more context" heading.

<u>Using the Collocation Feature:</u>

A collocation search is where you type a word or phrase into the WORDS box and a word or wildcard into the COLLOCATES box; good for discovering collocates that aren't necessarily adjacent

Remember to specify whether you want the collocate to come before or after the word (and how many words you want intervening) using the drop-down number bars to the right of the COLLOCATES box

When doing a collocation search you may also apply basic search techniques such as specifying the part of speech of a word, or specifying the wildcard as a certain part of speech.

When typing a wildcard into the COLLOCATES box, MINIMUM will automatically change to Mutual Info: 10. Change the 10 to 3. When not using the collocates box, change MINIMUM back to Frequency.

APPENDIX C

Error Correction Form

Name: _____
Date: _____

Error Correction Form

A. Re-write the sentence containing the error:

B. Indicate the search you conducted by filling out each line below. (You do not need to write down unsuccessful searches; only write down the search that ended up helping you).

___ List   ___ KWIC

WORD(S) _____

Did you take a sample? ___ Yes   ___ No

Which text types did you choose:
___ Spoken   ___ Fiction   ____ Magazine   ____ Newspaper   ____ Academic

C. Copy down one of the lines from the corpus that helped you fix your error:

D. Describe why the error is wrong and what you did to fix it:

E. Rewrite the sentence with your correction:

APPENDIX D

Attitude Survey

Corpus Experience Survey

<u>Description:</u> This survey is designed to collect information about students' corpus use experience.

<u>Instructions:</u> Read each question carefully and select the answer that best describes you.

Your insight is extremely valuable and your honesty is appreciated. The more honest you are, the more this survey will be able to help us improve corpus instruction in the future.

An AUS iLearn technician has assured us that each of your surveys will be ANONYMOUS.

Thank you for taking this survey!

1.   How good do you consider your computer skills to be?
    ___ Better than most
    ___ Average
    ___ Weaker than most

2.   Was this your first time using a corpus?
    ___ Yes
    ___ No

3.   Did you attend both of the training sessions?
    ___ Yes
    ___ No

4.   **Writing 001**

    I think the corpus is a useful tool for students wanting to improve their English.
    ___ Strongly agree
    ___ Agree
    ___ Not sure
    ___ Disagree
    ___ Strongly disagree

    English 207

    The corpus provides answers to my questions
    ___ Most of the time
    ___ Sometimes
    ___ Rarely

5. I think the corpus is a useful tool for me.
    ___ Strongly agree
    ___ Agree
    ___ Not sure
    ___ Disagree
    ___ Strongly disagree

6. I feel confident in my ability to use the corpus.
    ___ Strongly agree
    ___ Agree
    ___ Not sure
    ___ Disagree
    ___ Strongly disagree

7. I think I will continue to use the corpus.
    ___ Strongly agree
    ___ Agree
    ___ Not sure
    ___ Disagree
    ___ Strongly disagree

8. Please include any other comments about your experience using the corpus.

9. I am conducting interviews with interested participants to gain further insight into students' experience using the corpus. If you would like to be interviewed, please leave your information below. Interviews will last approximately 15 minutes.
    Name:
    Email:
    Preferred time:

APPENDIX E

Follow-Up Survey

Follow-Up Survey

Welcome to the survey!

This very short survey is designed to follow-up on your use of the corpus. You will see some questions with blue text. You can hover over the blue text and it will provide more information.

Thanks for your time!

1. What is your native language?
___ Arabic
___ English
___ Farsi
___ Hindi
___ Malayalam
___ Russian
___ Tagalog
___ Urdu
___ My native language is not listed here
___ I speak more than one language (please explain): _____

2. Regarding English, are you a
___ Native speaker
___ Near-native speaker
___ Non-native speaker

3. What is your opinion on the following statement? "I would like to receive more training in how to use a corpus."
___ Strongly agree
___ Agree
___ Not sure
___ Disagree
___ Strongly disagree

4. Have you used the corpus since you were introduced to it six weeks ago?
(If yes, students were directed to questions 5-7. If no, students were directed to question 8).
___ Yes
___ No

5. What factors led you to continue using the corpus?

_____

_____

_____

6. How much have you used the corpus since you were introduced to it six weeks ago?
___ Five times or less
___ More than five times

7. On a scale of 1-5 (1 = not at all, 5 = huge improvement), how much do you think your corpus skills have improved?
___ 1
___ 2
___ 3
___ 4
___ 5

8. What factors deterred you from using the corpus?

_____

_____

_____

_____

VITA

Shireen Palmer Baghestani was born in Boulder, Colorado, on November 5, 1986. She graduated from Niwot High School in 2005 with an International Baccalaureate diploma. For her undergraduate studies, she attended Miami University in Oxford, Ohio on a full ride academic scholarship. She graduated in 2009, summa cum laude and with University Honors with Distinction, with a B.A. in Linguistics and a minor in Middle Eastern and Islamic Studies.

Shireen then moved to the United Arab Emirates where she enrolled in the MA TESOL program at the American University of Sharjah. While pursuing her graduate studies, Shireen worked as a Graduate Assistant in the Department of English. She also worked as a substitute teacher and tutor at the AUS Achievement Academy Bridge Program. In addition, she worked as a research assistant on a project to uncover faculty and academic staff views of iLearn (Blackboard) at AUS. Shireen was awarded her MA TESOL degree on June 14, 2011.