Brigham Young University BYU ScholarsArchive

# Teaching Vocabulary Through Data-Driven Learning 

Erin Margaret Shaw<br>Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd
Part of the Linguistics Commons

## BYU ScholarsArchive Citation

Shaw, Erin Margaret, "Teaching Vocabulary Through Data-Driven Learning" (2011). All Theses and Dissertations. 3024.
https://scholarsarchive.byu.edu/etd/3024

# Teaching Vocabulary Through Data-driven Learning 

Erin M. Shaw

# A Project submitted to the faculty of Brigham Young University in partial fulfillment of the requirements for the degree of <br> Master of Arts 

Dee Gardner, Chair William Eggington<br>Neil J Anderson Mark Davies<br>Department of Linguistics and English Language<br>Brigham Young University<br>June 2011

Copyright © 2011 Erin M. Shaw
All Rights Reserved


#### Abstract

Teaching Vocabulary Through Data-driven Learning

Erin M. Shaw Department of Linguistics and English Language, BYU Master of Arts

The purpose of this master's project was to write a resource book that demonstrates how teachers can use data-driven learning methods to teach vocabulary. First, a brief overview of corpus linguistics, data-driven learning, and the corpus used in this book (COCA) is given. Then, the book presents different aspects of vocabulary learning in the context of a corpus. Topics included are frequency knowledge, part of speech knowledge, morphological knowledge, synonym knowledge, collocational knowledge, and register knowledge with a chapter on each topic. For each aspect of vocabulary learning there is a section that introduces the topic to the teacher, followed by instructions on performing topic related searches in the corpus. Each chapter also includes examples and ideas for application to the vocabulary classroom. Additional chapters provide information on individual language learning, and an evaluation of the project.

The goal of this project was to provide teachers with specific knowledge of vocabulary and corpus-linguistics to be able to teach less-frequently addressed aspects of vocabulary instruction and to encourage more use of corpora in the language classroom. It is hoped that after reading this book, teachers will be able to improve their vocabulary teaching and ability to use the Corpus of Contemporary American English and DDL methods in the ESL/EFL classroom. The evaluation of this project will consist of teacher reviews of the book after reading. Specifically, the questionnaire addresses readers' feelings of increased knowledge and understanding of these areas and desire to use them in the classroom.


Keywords: COCA, data-driven learning, vocabulary teaching, metalinguistic awareness, DDL

## ACKNOWLEDGEMENTS

I am deeply grateful to all of my professors who have helped further my education and encouraged me along the way. I am especially grateful for the members of my committee for their corrections and feedback to make this document better than I could on my own. I am particularly grateful for Mark Davies for his help and encouragement ever since I was an undergraduate student with an interest in corpora. I also owe many thanks to Dee Gardner for his patience, support, compassion, mentorship and always believing in me. I also thank Neil Anderson and William Eggington for their counsel and advice along the way. I am also truly appreciative of all my Linguistics and TESOL professors who have inspired me, pushed me and given me scholarly and Christian examples to follow. I would also like to thank my family and friends for their support, editing and feedback, and my students for their willingness to try new things. Most of all I would like to thank my Heavenly Father, who provided me the push to keep going when I needed it, the calm when I felt stress, and the inspiration for all of the following material.

## Table of Contents

Table of Contents ..... iv
List of Tables ..... vii
List of Figures ..... viii
Preface ..... X
Chapter 1: Corpus Linguistics in Language Education ..... 1
Introduction ..... 1
Data-driven learning ..... 3
Possible benefits and limitations ..... 4
Lack of DDL in the classroom ..... 8
DDL and Vocabulary Instruction ..... 9
The Corpus of Contemporary American English ..... 12
Summary ..... 14
Chapter 2: Frequency Knowledge and Example Sentences ..... 15
Introduction ..... 15
Using the corpus ..... 18
Applications and sample lessons ..... 21
Chapter 3: Part of Speech Knowledge ..... 23
Introduction ..... 23
Using the corpus ..... 25
Applications and sample lessons ..... 29
Chapter Appendix ..... 33
Activity: Identifying Parts of Speech ..... 33
Chapter 4: Morphological Knowledge ..... 36
Introduction ..... 36
Using the corpus ..... 39
Applications and sample lessons ..... 41
Chapter Appendix: ..... 46
Lesson Plan: Derivational Morphemes Part I ..... 46
Lesson Plan: Derivational Morphemes Part II (Greek and Latin Roots) ..... 53
Chapter 5: Synonym Knowledge ..... 59
Introduction ..... 59
Using the corpus ..... 61
Applications and sample lessons ..... 64
Chapter 6: Collocational Knowledge ..... 68
Introduction ..... 68
Idioms and other phrases ..... 69
Mutual information ..... 71
Using the corpus ..... 72
Applications and sample lessons ..... 75
Chapter Appendix: ..... 80
Lesson Plan: Learning Academic Vocabulary With Data-driven Learning ..... 80
Chapter 7: Register Knowledge ..... 88
Introduction ..... 88
Using the corpus ..... 90
Applications and sample lessons ..... 95
Chapter Appendix: ..... 99
Lesson Plan: Learning about Register (Academic English) ..... 99
Chapter 8: Extension (Autonomous Learning) ..... 105
Individual Language Learning ..... 105
Case studies ..... 106
Troubleshooting COCA searches ..... 110
Other Corpus-based Vocabulary Resources. ..... 112
Chapter 9: Teacher Feedback ..... 115
Chapter Appendix: ..... 119
Evaluation ..... 119
References ..... 121

## List of Tables

Table 1. Explanations of What it Means to Know a Word ..... 11
Table 2. Breaking Down Vocabulary Words into Parts (Incomplete). ..... 42
Table 3. Breaking Down Vocabulary Words into Parts (Partially complete) ..... 43
Table 4. Breaking Down Vocabulary Words into Parts (Complete). ..... 44
Table 5. Synonym Chains and Examples of Academic Vocabulary ..... 65
Table 6. Collocational searches in The Corpus of Contemporary American English. ..... 74
Table 7. Word Meaning and Collocates from News Listening Passage ..... 77
Table 8. Vocabulary Words in Collocational Phrases and COCA searches. ..... 78

## List of Figures

Figure 1. Registering for COCA. ..... 18
Figure 2. Frequency search of walk the dog. ..... 19
Figure 3. Concordance lines of walk the dog. ..... 19
Figure 4. Help screen and starting point for COCA tours ..... 20
Figure 5. Question marks to access more help information. ..... 20
Figure 6. Key Word in Context radio button. ..... 25
Figure 7. Key for colors used in KWIC radio button. ..... 25
Figure 8. KWIC of rock out. ..... 26
Figure 9. Wild card search for a phrase. ..... 27
Figure 10. Part of speech tag list. ..... 28
Figure 11. Part of speech wildcard phrase search. ..... 28
Figure 12. How to turn on part of speech tags on words ..... 29
Figure 13. Frequency search showing part of speech tags ..... 29
Figure 14. Wild card search for grammatical forms. ..... 40
Figure 15. Lemmatized search of study. ..... 40
Figure 16. Lemmatized phrase search of run errands. ..... 41
Figure 17. Simple synonym search of beautiful. ..... 62
Figure 18. Lemmatized synonym search for inquire. ..... 63
Figure 19. Synonym phrase search. ..... 63
Figure 20. Word web of synonyms for comply. ..... 67
Figure 21. Simple wildcard collocate search for open ..... 72
Figure 22. Part of speech collocate search. ..... 73
Figure 23. Search limits for collocate searches. ..... 73
Figure 24. Chart search comparing neglect across registers. ..... 91
Figure 25. List search and showing register frequency ..... 91
Figure 26. KWIC of however in the academic register. ..... 92
Figure 27. Dunk in the sports magazine sub-register. ..... 92
Figure 28. Selecting registers to compare under the sections heading. ..... 93
Figure 29. Comparing the collocates of two sub-registers. ..... 94
Figure 30. Wildcard phrase search comparing two registers. ..... 94
Figure 31. Synonym search compared in two registers ..... 97
Figure 32. Collocate search compared across two registers. ..... 97
Figure 33. Selecting frequency or relevance settings. ..... 111

## Preface

This book is meant to be an introduction to principles of vocabulary teaching and teaching with corpus linguistics, i.e. data-driven learning. Because it is an introduction, it is not comprehensive in covering all the complex issues involved with these topics. Rather, it gives the reader possibilities and limitations. The goal of this book is to introduce principles and applications of vocabulary teaching and data-driven learning to encourage language teachers to explore them further on their own and use them in the classroom. Hopefully, readers will come away from it with an increased understanding of both vocabulary and datadriven learning, as well as a desire to incorporate both into their classroom more.

Chapter 1 begins with an introduction to the role that corpus linguistics and datadriven learning have played in language education. It also includes a brief introduction to the corpus used in this book, The Corpus of Contemporary American English (COCA), by Mark Davies (2008-). Chapters 2-7 each introduce a principle of vocabulary teaching that can be taught through DDL, along with guided instructions on performing corpus searches and applications to the language classroom. Chapter 8 demonstrates additional applications of these classroom principles to individual learning, and Chapter 9 includes teacher viewpoints and attitudes about this book.

Chapters 2-7 include a simple overview of each of the vocabulary topics presented in this volume in the "Introduction" section and why they are important for vocabulary learning. Topics covered are: frequency and example sentences, part of speech knowledge, morphological knowledge, synonym knowledge, collocate knowledge, and register knowledge. These subjects were chosen because of their essential nature in developing
metalinguistic awareness of words and because of the way the corpus interface allows for easy exploration.

In order for teachers to be able to create DDL activities beyond what are presented in this book, it is essential for them to understand how searches and activities were created from COCA, the corpus used. Consequently, Chapters 2-7 also give detailed instructions on performing searches related to the vocabulary constructs presented in the "Using the corpus" sections. Each chapter builds upon the previous chapter for both vocabulary principles, and also corpus search terminology and instructions. In this way, neither teachers nor learners are forced to learn all of the corpus features at once; individual chapters can be examined with understanding of previous chapters. Included for reference and clarity are: screen shots of sample searches, quotation marks to show the exact search string entered into the corpus (e.g., "walk the *") or features of the corpus (e.g., "list" button), and italics to show a general usage of the word being queried (e.g., hard up). Teachers may use the instructions in each "Using the corpus" section for themselves or adapt them for students. More technologically savvy teachers might even create a screen recording video to demonstrate to students how they want them to use the corpus before activities.

Finally, Chapters 2-7 also include a section titled, "Applications and sample lessons," which aims to demonstrate and generate ideas of how to apply DDL to the classroom. Activities vary in complexity and length to show how DDL can occupy whole lessons or small activities depending on the class size and desired outcomes. Appendices in some chapters give additional lesson plans to introduce and teach concepts on a larger level.

Based on the author's personal experience and student interviews, if students are too preoccupied with learning the language of the corpus interface, and the linguistic knowledge
to make it work, DDL activities can become too time consuming to be useful. Therefore, incremental introduction and use of the corpus in class can be more effective. Individual teachers can also be aware of new information overload by adapting the amount of information shared and how DDL and vocabulary concepts are presented to their classes. Previous experience by the author has shown that many students need several practice sessions with one feature of the corpus before they feel comfortable using it on their own.

Though learner autonomy in general is encouraged in DDL, the emphasis of the DDL methods in this book is on in-class learning, where the teacher is a resource and guides the students through patterns of noticing and awareness. In order for this to happen, teachers should be familiar with the corpus features, help students develop intuitions and metalinguistic awareness, and encourage intellectual curiosity of their students to search for patterns on their own. After having performed searches and in class DDL, students might be able to apply the processes practiced in their individual study. Chapter 8 gives examples of DDL in one-on-one settings and suggestions for adaptations of DDL for different students. However, even in individual DDL students often find it helpful to consult native-like English speakers in order to differentiate between exceptions found in the corpus, add pragmatic information, and interpret difficult and confusing results.

Finally, the author would like to state that, based on her teaching experience, these methods can help students and teachers experience language learning in a whole new, exciting way. Hopefully, this book is able to communicate and accomplish that.

## Chapter 1: Corpus Linguistics in Language Education

## Introduction

Corpus linguistics is not rationalistic; it is empirical. It thrives on data to analyze and discover what language speakers do. Large bodies of text (corpora) reveal these patterns in words, grammar, and discourse. When computers aid this process, those texts--either spoken or written--can be handled rather quickly, especially if they are tagged for parts of speech or specific information. Corpora allow access to authentic data; they reveal knowledge about the language that is not intuitive and show frequency patterns of words and grammar constructions. Such patterns can be used to improve language materials or to directly teach learners. As J. Flowerdew (2009) puts it, "no dictionary or grammar is able to fully describe the language" (p. 329) so corpus linguistics fills in the gaps by helping us describe missing pieces that students might not be learning traditionally from other data.

The boon of modern technology has influenced linguistic research and language teaching by allowing for methods of research not possible without it. Though corpus linguistics existed before Chomsky and rationalism dominated linguistic theory, it was small and underground for several years (McEnery \& Wilson, 2001). Yet, corpus linguistics never completely died. With the advent of faster and more powerful computers in the eighties, along with other turns in linguistic theory, corpus linguistics came alive stronger than it was before as a viable linguistic methodology. Since that time, it has only continued to grow in use and applications. Because corpus linguistics is not a branch of linguistics, but a methodology, it can be applied to every branch of linguistics (McEnery \& Wilson, 2001).

The most recent branch of linguistics to benefit from applications of corpus linguistics is language teaching. Romer (2009) claims that many corpus linguists believe that "corpus
linguistics can make a difference for language learning and teaching and that it has an immense potential to improve pedagogical practice" (p. 84). Such thinking is similar to that of pioneers of corpus linguistics like Tim Johns (1991) and John Sinclair (1988), who have long advocated using corpus linguistics in language teaching. Much of Sinclair's work was dedicated to the COBUILD project, the first corpus-based dictionary. His work with that dictionary had massive pedagogical implications, and forever changed learner dictionaries (Moon, 2007). Tim Johns, also known as "Mr. DDL," was a strong advocate of data-driven learning, or letting learners induce patterns from corpora, in the language classroom. Data-driven learning (DDL) includes both deductive and inductive processes (Hadley, 2002); however the primary focus has traditionally been looking at examples to induce rules or patterns (Johns, 1991).

Bennett (2010) describes three ways corpora can be incorporated into language teaching. These include: corpus-influenced materials, corpus-cited texts, and corpus-designed activities. Corpus-influenced materials include textbooks and classroom materials that are not just based on traditional grammar rules, but patterns and frequency information obtained from corpora. Corpus-cited texts are dictionaries and grammar books based on corpus data. Corpus-designed activities are those in which the learner actually sees or is involved with the actual data in some way, or essentially DDL. Language education is continually influenced by corpus linguistics at a rapid pace in each of these three categories. A simple Internet search of corpora and language teaching will reveal many journal articles and books published within the last ten years which deal with corpus applications to language teaching in some of the aforementioned ways. For a small sampling see Aijmer, 2009; Campoy-Cubillo, Belles-Fortuno, \& Gea-Valor, 2010; Ghadessy, Henry \& Roseberyy, 2001; O’Keeffe, McCarthy, \& Carter, 2007; Sinclair, 2004.

Since Johns and Sinclair first introduced their ideas, many others have expanded on their work and the popularity of corpora in language teaching has continued to increase (Sinclair, 2004). However, there is a gap between their research and the classroom, as evidenced by teachers' and students' reticence to use these materials (Aijmer, 2009; Pérez-Paredes \& AlcarazCalero, 2009). Among the many hypotheses to explain this gap is the conclusion that teachers do not have enough awareness of what a corpus is, nor are they informed about corpora in teacher education programs (McCarthy, 2008). Teachers themselves must have a basic understanding of data-driven learning and corpora before they can effectively demonstrate to their students how they can help and how they can incorporate them into the curriculum (Sinclair, 2004). Therefore, one purpose of this book is to add to teachers' knowledge of corpora in language teaching, specifically DDL, in a small attempt to bridge the gap.

## Data-driven learning

DDL is a term coined by Tim Johns (1991) to describe language learners acting as "language detectives." In the strictest sense of the term, it is a way of using corpus linguistics in teaching by exposing learners to data and having them find the rules and patterns from example concordance lines, or lines of text from a corpus that focus around a single word. Johns described his method to learning as "...the perception that 'research is too serious to be left to the researchers': that the language-learner is also, essentially, a research worker whose learning needs to be driven by access to linguistic data-hence the term 'data-driven learning' (DDL) to describe the approach" (Johns, 1991, p. 2). Johns' use of DDL was typically with advanced level language students, often highly educated post-graduates, who focused on both complex and minute rules of language. He primarily used DDL as an inductive process, where often times the teacher did not even know the answer or the patterns that the student would discover (1991).

While such discovery based methods--with unknown results and a narrow audience of highly proficient language learners--were successful for Johns, such methods and applications of DDL might not work well for a broader audience of language learners. Thus, DDL has moved into mainstream teaching, through expansion of Johns' original approach. Current DDL practices vary and might include rule discovery and concordance lines like Johns, but also teacher guided searches of known rules with other corpus features, and different kinds of corpora such as learner corpora (Aijmer, 2009; Balunda, 2009; Basanta \& Martin, 2007; Boulton, 2009b, 2010; Clifton \& Phillips, 2006; Cresswell, 2007; Hadley, 2002). Furthermore, DDL is no longer restricted to inductive learning. Students can also learn rules and look at real language samples to deduce further examples based on observed patterns and teacher explanations (Chujo, Anthony, \& Oghigian, 2009). In addition, DDL has traditionally focused on concordance lines, or incomplete sentences that focus on a common word, but new research and corpora have also used lines of texts with expanded context to increase learning possibilities (Boulton, 2009b; Davies, 2008-). Depending on the corpus interface, on-line corpora often offer other ways of viewing the corpus data as well, some of which will be discussed in later chapters. Therefore, with these expansions on Johns' original work, DDL is no longer looking only at concordance lines for patterns, though those methods are still used. This book adapts an expanded view of DDL and sees it as any sort of activity that looks at chunks of real language data from corpora to aid learning.

## Possible benefits and limitations

The authentic language of DDL is one of the greatest benefits that is cited in its favor (Clifton \& Phillips, 2006; Romer, 2008). In this way, DDL allows teachers and students to study naturally occurring language for grammatical patterns, word usage, semantic and pragmatic features, and textual discourse (L. Flowerdew, 2009). Additionally, authentic language allows for
language topics that otherwise might not occur in a textbook or from the examples created in teachers' minds, to be used in vocabulary and grammar discussions. Finally, since teachers cannot change the language students encounter in English speaking environments, DDL, in providing authentic examples and encouraging "noticing," or "awareness-raising," helps students of all levels develop skills needed to deal with the language around them (Romer, 2008).

Another possible benefit of DDL is that it allows for more potential learner autonomy in the classroom, with less reliance on the teacher. Such is related to the premise Johns supported of the student as a researcher. This autonomy, as well as inductive and deductive reasoning methods, gives opportunities for the development of cognitive skills (Boulton, 2009b). Also, DDL's hands-on approach to learning provides students who struggle with traditional methods an alternate approach. With greater autonomy, students can also use DDL techniques to answer their own questions about language, as well as to become more independent language learners (Hunston, 2002).

On the other hand, some of these same benefits of DDL may be seen as possible limitations, especially because they have yet to be proven. Though there have been many studies that have shown increased learner benefits from DDL instruction, not all of them have proven empirically that DDL works better than traditional methods; some have been primarily based on learner's perceptions of improvement (Cresswell, 2007). As Boulton (2009b) puts it, "More empirical studies are needed to indicate different conditions for use of DDL - for what types of learners, what minimum resources, what language points, how it can be integrated with other techniques, and so on" (p.51). Therefore, the need for more empirical proof is a limitation of DDL.

Other potential arguments against DDL are the underlying assumptions it has about second language acquisition (Johansson, 2009). These include complex issues such as implicit and explicit learning, learner awareness through consciousness raising, and arguments against authentic language. Certain beliefs concerning the role of these topics in language learning might lessen DDL's credibility as a teaching method for some individuals. For example, authentic language has great supporters (Braun, 2005), but others are staunchly against it, particularly in the corpus sense. Widdowson (1998) summarizes his arguments against it as follows:

For the promoters, authentic language means the actual linguistic text that people produce and that can be collected and analysed by computer. This analysis yields many a fascinating fact about frequency of occurrence and patterns of co-occurrence. But what it cannot yield is information about how the texts thus analysed interacted with contextual conditions to realise discourse. It cannot tell about the discourse process whereby pragmatic meaning is appropriately achieved: It can only record the overt attested product of that process, its textual trace. (p. 712).

Therefore, DDL can be limited by authentic language because it can be difficult to fully understand without being an insider in the context (Widdowson, 1998). Additional assumptions made by DDL supporters are that students are motivated to remember knowledge they have worked to discover on their own and that as students study examples and forms of large quantities of data, previously unnoticed patterns will appear. Neither one of these claims have been proven specifically (Hunston, 2002).

In general, DDL supporters do not hide the lack of conclusive evidence or agreement behind these possible limitations, but consider them areas of further exploration and continuing research (Boulton, 2009a). Notwithstanding, it is important to note that eliminating these assumptions does not remove all of the benefits of DDL, because it has never been proven to not be an effective method. Many studies have shown its usefulness as a teaching method in language classrooms (Aijmer, 2009; Balunda, 2009; Boulton, 2009a, 2009b, 2010; Chujo, et al., 2009; Cobb, 1997; Cresswell, 2007; Fan, 2009; J. Flowerdew, 2009; L. Flowerdew, 2009; Hadley, 2002; Henry, 2007). For further review of pros and cons of DDL, consider reading Boulton (2010), L. Flowerdew (2009) and Romer (2008).

Finally, one last concern raised about DDL is the skill level when it is used. One common argument against using DDL in the classroom is that it is only applicable to advanced learners, as in how Johns used it (Boulton, 2009a). However, recent research has shown that traditional instruction is not more effective than DDL at lower skill levels (Boulton, 2009b; Hadley, 2002). Effective DDL at low-levels is true even in monolingual contexts where children have been shown to be able to manipulate native-language corpus data to their benefit (Sealey \& Thompson, 2007). Boulton (2009b; 2010), perhaps the most avid supporter of DDL at lowerlevels, further explains that because DDL allows for different learning styles, struggling students benefit from alternate methods of instructions. He says, "Perhaps DDL might be appropriate for 'less successful' learners as an alternative to traditional methods which, by definition, have not worked in their case" (Boulton, 2009b, p. 39). This comment supports the view that DDL is not a replacement for traditional teaching methods; instead, it is another arrow in a teacher's quiver to help his students. However, though there may be a place for low beginners in DDL, activities and
lessons in this book will concentrate on high beginning to advanced levels only for greater ease of adaptability and flexibility.

## Lack of DDL in the classroom

Though there is growing support in the linguistics community for DDL and corpus-based language teaching, it still has not become common to the language classroom. Several possible reasons include lack of awareness concerning the existence of DDL, lack of understanding of DDL processes, unwillingness to try new methods, not knowing how to integrate DDL into the language classroom, fears of the technology, fears of not being able to interpret corpora, DDL materials being too detailed and rule specific for teachers, lack of DDL resources to draw from or DDL materials and methods not meeting teachers' needs (Boulton, 2009a; Hunston, 2002; McCarthy, 2008; Romer, 2009). Depending on the language-learning context, any one of these reasons could be true. In addition to these reasons, teachers might choose not to use DDL because of student pressure and fears. For learners, fears often include new methods, different personal learning styles, computer use, and low face-validity (Boulton, 2010). Moreover, all of these possible reasons for not using DDL are in addition to the arguments and limitations about using DDL in general (e.g. Flowerdew, 2009).

Though the aforementioned reasons are all valid reasons why DDL is not used, McCarthy (2008) specifically focuses on one of them, teacher's lack of education concerning corpora in language teaching. He says, "In my travels across several continents, my common experience is that teachers have heard of corpora, but they are not quite sure what they are..." (McCarthy, 2008, p. 563). He continues to explain that much of literature written about corpora in language teaching and DDL has been done by researchers, academics, who decide this is what teachers need. He purports that corpus-based methods will not catch on until the teacher becomes the
informed consumer, knowing how to evaluate corpora and how to effectively use them--all points that Boulton (2009a) and Romer (2009) agree with. In sum, researchers agree that in order for DDL to be used more in the classroom, language teachers need more knowledge and awareness concerning it and corpora in general.

Out of all the possible reasons for not using DDL, a lack of teacher awareness is probably the easiest to resolve and this book hopefully provides a partial solution. Within this volume are introductions to basic principles, instructions, and applications to apply DDL in the classroom. Hopefully, teachers will be able to benefit from the information presented and will begin to overcome both their fears and their students' fears concerning DDL through greater understanding. By framing DDL through the medium of vocabulary teaching, readers will hopefully be able to understand more of the what and how of DDL, which can later be applied to other language areas such as grammar teaching. Vocabulary teaching also has some natural overlap with principles of corpus linguistics that permit easy applications of DDL. Consequently, the next section will introduce some of these connections and set the foundation for later chapters, which will expand on some of these specific principles.

## DDL and vocabulary instruction

Corpus linguistics has already influenced much of what is known about vocabulary and vocabulary instruction (Nation, 2001; Schmitt, 2000). Because of the nature of corpora, there is a natural, close relationship between corpus analysis and vocabulary knowledge (Read, 2010). The developments of corpora and corpus materials specifically have affected what we know about vocabulary issues such as grammar, idioms and other phrases, choosing what words to teach, and knowing what a word means. Therefore, applying DDL to vocabulary instruction almost seems
intuitive. DDL methods in the classroom allow some of these issues that might not otherwise be addressed effectively to be explored in new ways.

Vocabulary is key to language learning, and needs to be learned effectively. Unfortunately, teachers often overestimate just how much vocabulary students really understand, and therefore fail to teach at a level of comprehensible input (Folse, 2004). Recently, the focus on vocabulary acquisition and knowledge in language teaching has increased, and our understanding of the issues involved in learning new words has deepened. Particularly, we now know that vocabulary knowledge and learning is a lot more than learning the definition of a word (Nation, 2001). Metalinguistic awareness is a term used to describe the knowledge of rules about language, parts of language, and how language works, or even a simple self-recognition of one's own language and the forms being used (Koda, 2000). Therefore, metalinguistic awareness includes the idea that knowing a word is more than knowing its dictionary definition, but also includes knowing the word's spelling, morphology, parts of speech, pronunciation, variant meanings, collocations, specific uses, and register related contexts of use (Nation, 2001). In addition to these, J. Flowerdew (2009) adds colligation, semantic preference, and semantic prosody. Corpus linguistics can offer valuable insights for many of these aspects of knowing a word, particularly because we now understand that these issues are affected by the vocabulary word as well as the surrounding context. Table 1 gives an overview and examples of theses issues.

Table 1

Explanations of what it Means to Know a Word

| Aspect of word <br> definition | What it means | Example |
| :---: | :---: | :---: |
| Concept | Understanding the background ideas of <br> the word. | Being able to understand that a hat is a <br> large object placed in the head. |
| Spelling | The sequence of letters in order for the <br> orthographic form. | Resplendent: R-E-S-P-L-E-N-D-E-N-T |
| Morphology | How affixes attach to base forms of |  |
| words. |  |  |$\quad$ Playfully: Play+ful+ly

Corpora can be helpful in teaching these aspects of metalinguistic awareness and demonstrating their usefulness. Some aspects of metalinguistic awareness, like register and collocation, are more accessible through corpora than other ways of learning words, such as dictionaries. Other aspects of metalinguistic awareness such as morphology, part of speech, spelling and specific uses can be learned through other sources like dictionaries, but corpora can provide additional knowledge. Corpora can also be used to teach some aspects of vocabulary learning not included as a part of metalinguistic awareness, like synonymy and frequency, but are still essential to word knowledge and interconnected with other aspects of metalinguistic
awareness. Teachers who are aware of this spectrum of vocabulary issues can aid their students in comprehending the full meaning of words and practicing them in the classroom.

Metalinguistic awareness and vocabulary teaching are greatly informed by corpus linguistics. Therefore, using DDL in vocabulary teaching fits logically. Their close relationship provides great opportunities for learning more about vocabulary and how to apply DDL in the language classroom. Both will be explored together in the following chapters through specific aspects of vocabulary learning. These include: frequency knowledge and example sentences, part of speech knowledge, morphological knowledge, synonym knowledge, collocational knowledge, and register knowledge. Before these ideas are presented, a brief introduction to the corpus used in this book will be given and how it facilitates both DDL and vocabulary learning.

## The Corpus of Contemporary American English

In order to provide a useful tool for DDL, an appropriate corpus must be selected, particularly one that enables study of metalinguistic awareness and is accessible enough for DDL. All activities in this book are based on The Corpus of Contemporary American English (COCA) (Davies, 2008-). Principle reasons for using COCA are its large size, diverse structure, focus on American English, user-friendly interface, easy accessibility and ability to show features of vocabulary learning. While other free corpora exist, COCA, created by Mark Davies in 2008, is the largest, free English corpus available and has significant advantages over other free corpora (Davies, 2009).

In terms of vocabulary study, COCA has definite advantages. First, COCA currently has over 425+ million words, with 20 million words being added on a yearly basis. Such a large size gives a sufficient patterning of English lexis and grammar without misrepresenting outliers as the norm (Davies, 2009). In terms of vocabulary, this means that the corpus will give an appropriate
picture of word frequency in terms of how they are really used. Second, COCA has the benefit of being a balanced corpus in terms of register, or type of language being used. The corpus is balanced equally between its five registers of spoken, news, academic, fiction and magazine (Davies, 2009). Therefore, users need not fear that the examples they find are only how people write in newspapers, or only how they speak. This gives students a more realistic interpretation of how and where words are used. Third, COCA is based only on American English, ensuring a consistent and constant reference point for language. This ensures that for students in the United States, vocabulary is appropriate to the learning context. Though not all language in the corpus may be "proper" English, all of the sources in COCA are published or broadcasted by American corporations and maintain some standard of acceptability (Davies, 2009). Fourth, the interface of COCA allows for easy access to non-linguists without them requiring to know specific part of speech tags or codes to perform searches. COCA also includes help guides and tours to learn how to use its features, whether for vocabulary or for grammatical searches. In addition, COCA is free and publicly accessible on the internet at www.americancorpus.org. Users only need to register an email address.

Finally, COCA's unique interface allows for features of metalinguistic awareness to easily be analyzed. The corpus is already tagged for part of speech and register information, and easily searches for collocates, synonyms, sub-registers, overall frequency and frequency over time. For users interested in more than vocabulary, COCA also allows users to search for grammatical forms, and patterns. The variety of search options in the corpus facilitates vocabulary study well. Because COCA is free and accessible to all those with an internet connection, it allows for DDL opportunities worldwide.

For a complete review of the features of COCA see Davies (2009) and Davies (2010) as well as the many tutorials and introductions available for novice users on the COCA website.

## Summary

This chapter has attempted to introduce the background of corpora in language teaching, a specific application of corpora to language teaching through DDL, and the possible benefits and limitations of DDL. Also included were reasons why DDL is not present in many language classrooms, and how because of the close connections between vocabulary and corpora, DDL relates to vocabulary instruction. In addition, a basic introduction was given to COCA, and how its specific features can be used for DDL and vocabulary teaching. The remainder of the book will build on this information by presenting specific aspects of vocabulary that can be taught using DDL activities from COCA.

## Chapter 2: Frequency Knowledge and Example Sentences

## Introduction

Though not specifically listed as a feature of metalinguistic awareness, frequency knowledge also improves word understanding and is an important consideration in vocabulary teaching. J. Flowerdew's (2009) review points out, "Knowledge of ... relative frequencies can be helpful to language practitioners in deciding what items to teach and when to teach them" (p. 330). Therefore, frequency knowledge builds off of other aspects of metalinguistic awareness like register and specific uses to know what words are most useful, according to a specific context. Deciding what words to teach and in what order is then done through specialized word lists based on the most frequent words in the language context.

Ferris (2009) identifies three levels of vocabulary for students in academic contexts: general English, general academic English and English for their specific discipline. Many frequency lists of the English language have been produced based on corpora, and can be used to help teachers identify words for each of those categories. The General Service List of English Words (GSL) (West, 1953) is perhaps the most famous of such lists, focusing on general English. The GSL is also the oldest, and therefore the most outdated; however, it is still quite useful because it separates words with similar spellings into different meanings. Davies and Gardner (2010) produced more updated word lists of general English and published the 5,000 most frequent English words in A Frequency Dictionary of Contemporary American English. Additional lists of varying sizes from them are also available online at http://www.wordfrequency.info. Coxhead's (2000) Academic Word List provides teachers with a specialized frequency list of words that occur specifically in academic literature. Her word list is used worldwide in many intensive English programs (IEPs) to develop general academic

English. Examples of word lists created for specific disciplines include medical vocabulary for RA's, and engineering vocabulary for undergraduate engineers (Wang, Liang, \& Ge, 2008; Ward, 2009). Frequency-based word lists help teachers build a strong foundation of words to teach, as well as cater vocabulary teaching to the contexts where students will use the language. In their own classrooms, teachers can use corpora to select words from a unit or book to teach based on frequency of occurrence. By learning frequent words, students have a greater chance of fully understanding the words, particularly for receptive and productive use, because of increased exposure (Romer, 2008).

Recently, using corpora to create word vocabulary lists based on frequency has become a more commonly accepted practice in vocabulary teaching (Schmitt, 2000). However Schmitt also argues that frequency is not the only consideration in determining what vocabulary to teach. Though frequency is a large factor, he also lists as additional factors theme specific (or technical) vocabulary, words that students choose, as well as classroom management words (p. 144). Gardner (2008) also argues that frequency is important, but within a narrow theme or specific area as the best way for teachers to choose which words to teach. Fan (2009) adds that teachers should include frequent words, but also their collocates, and teach them together. Therefore, though teachers should not focus on frequency alone, they can and should use corpora to create specialized word lists for their classroom, informed by frequency, collocation, and lexical patterns.

English has a large vocabulary and wide variety of words to choose from, and frequency knowledge can help speakers determine which words to use in which contexts. By introducing frequency knowledge, teachers help students appreciate how words are actually used, and how to select useful words on their own to learn. Typically, English speakers will use more frequent and
less technical words in general settings. The opposite is true as well; for specialized topics and discussions, less-frequent words with more specific meanings are used more often than general words. In a classroom setting, a beginning student might be proud to learn a difficult word, like abysmal, but disappointed when he hardly ever hears it used by native speakers. Hence, teaching students how to identify frequent words in a corpus can help them study even more effectively on their own.

For DDL activities in the classroom, a simple frequency search is the most basic activity to do in COCA. A frequency search shows how many times the type (or specific string of characters) occurs in the 425+ million-word corpus. A high-frequency verb like drift occurs 8938 times as opposed to a meander, which only occurs 1032 times or a low-frequency verb like dawdle with 352 times (at the time of writing). All of the verbs carry similar meanings, though they differ greatly in frequency. Students who engage in the less-than-fluent reading strategy of looking up all unknown words in a reading passage, adding them to their vocabulary list and trying to learn all of them, would benefit from frequency knowledge. On the other hand, corpustrained students who are aware of frequency may try to remember the word dawdle to understand the reading passage it is found in, but not spend too much time memorizing it because of its relatively low frequency. Those same corpus-trained students might also recognize that drift is more frequent and has a higher probability of appearing in later passages, and therefore make more of an attempt to learn it.

Another basic feature of COCA that is performed simultaneously with frequency searches, is the ability to show example sentences. Corpus linguists often refer to sentences that center around one key word (or node word) as concordance lines, or Key Word In Context (KWIC) lines. Concordance lines are ideal for DDL because they can provide a myriad of
examples in a manageable way without including the full context. Example sentences can give students an idea of how the word fits in grammatically with other words as well as clues to meaning through surrounding words. Concordance lines are the authentic language for students to digest in DDL. Instead of teachers creating contrived examples for exercises or tests, they have the possibility of real-life examples presented in COCA. Even if they need to be modified by the teacher to suit teaching purposes, authentic example sentences are still worth the time required to generate them from the corpus because they add variety in structure and context.

## Using the corpus

First, after 15-20 queries, all users must register to use the corpus. This can be easily done with any email address by clicking on the help bar towards the middle right of the screen and selecting "Register for COCA" (Figure 1).

| Help / information / contact |
| :--- |
| Help / information / contact <br> $====================$ <br> * Contact us / feedback <br> * Citing the corpus <br> * Your corpus-based publications/presentations <br> $=====$ GENERAL $=======$ <br> Where do I start?? <br> Brief tour (general) <br> Brief tour (for non-linguists) <br> Introduction (general) <br> Credits / history <br> Reqister for COCA <br> $=====$ COMPARE COCA $=======$ <br> Overview <br> Compare to the American National Corpus <br> Compare to the British National Corpus <br> Compare to the Bank of English <br> Compare to the Oxford English Corpus <br> Why not just use Google? <br> Use as a monitor corpus |

Figure 1. Registering for COCA.

To perform a basic frequency search, simply insert the word or phrase into the "word" query box and press "search." This can be done for words or phrases. The default is already set on the "list" radio button that will give the frequency of the search string. A basic frequency
search for "walk the dog" is shown as an example in Figure 2. To restart any search or change the words queried, simply hit the "reset" button next to "search".
$\square$

Figure 2. Frequency search of walk the dog.

In order to see real-world examples (concordance lines) of the word in context, simply click on the results of the search shown in blue. For example, Figure 2 shows "walk the dog" with the frequency count of 78 to the right of it, both in blue. Clicking on either the phrase or the frequency count will yield example sentences. By sliding your cursor to the bottom of the screen, the examples will fill the screen. The results will look similar to Figure 3. This is also known as the KWIC display.
$\square$

Figure 3. Concordance lines of walk the dog.

Tours within the corpus interface introduce frequency and KWIC screens as well as other
features. For a more detailed tour of these or other features of COCA, visit www.americancorpus.org. Click on the small white information bar that reads
"help/information/contact" and click on "Where do I start?" (Figure 4). Clicking the question mark icon next to different features (Figure 5) will also give a more detailed description of the feature and sample searches of how they can be used.


Figure 4. Help screen and starting point for COCA tours.
DISPLAY
©LIST OCHART OKIC OCOMPARE
SEARCH STRING
WORD(S)
COLLOCATES
POS LIST
RANDOM $\quad$ SEARCH RESET
SECTIONS $\square$ SHOW

Figure 5. Question marks to access more help information.

## Applications and sample lessons

Even the most basic frequency searches and example sentences can be helpful in learning vocabulary in the classroom in a variety of ways. First, frequency can help narrow textbook vocabulary lists to the most common words. When presented with large lists of words or phrases, frequency can also help teachers and students prioritize what words are most important since there are often too many to tackle at once. Next, frequency can help identify the usefulness of unknown words students encounter. Often students are taught to look up unknown words in the dictionary. When they do this, they may find definitions, but not any information telling them if the words are practical to learn. Frequency can help students measure if a word is low, medium or high frequency based on guidelines given to them by their teacher. In addition, frequency can be used to check spelling, since misspelled words have extremely low frequency or do not appear in the corpus. Furthermore, classroom instruction can focus on guiding students to develop intuitions and make guesses about whether or not a phrase is acceptable or not in English, and then to follow up with a frequency search in COCA to see if it is found in the corpus. One small caveat to using COCA for spelling and acceptability issues is the possible fallibility of the corpus. As a consequence of being based on real data, the corpus preserves spelling or usage errors that native speakers make in speech or writing. Teachers can help students get a feel for the relative frequency of items in the corpus by judging an item that appears only a few times out of more than 400 million words as not necessarily correct, widely-used in English, or the absolute truth.

Consider the following examples of using frequency and example sentences in the classroom. In Mr. Erickson's intermediate English class, students are assigned a vocabulary list of phrasal verbs to learn each week. His students are given the following phrasal verbs as part of
classroom expressions: mess around, use up, call on and wrap up. Mr. Erickson gives examples of what the phrasal verbs mean when he introduces them in class. He feels like his students need more examples because they struggle with how to use the phrasal verbs. He shows them in class how to look them up in COCA and find example sentences. For a homework assignment, Mr. Erickson has students look up and write down three examples of each phrasal verb from COCA. As a follow up, he has students share their example sentences in pairs to make sure they have good examples and understand the phrasal verbs.

In a high-beginning writing class, Mrs. Michaels has students use the corpus to check spelling and non-English word errors on their 30-minute essay paragraphs. Students are not allowed to use a spell checker while writing. Before she even corrects their papers, Mrs. Michaels then asks her students to search for the words they think may not be spelled correctly, or that they are unsure of in COCA. If the frequency of the word is below 10 , she asks them to circle it so she can help them identify what they perceive as errors later. She also asks them to keep a list of all circled words on a separate sheet of paper to keep track of their errors. Since students in her high-beginning class are not likely to use words in English that have a low frequency in the corpus, this proves to be a good strategy. After correcting their essays, Mrs. Michaels shows students the correct spelling and frequency of commonly misspelled words in the corpus.

## Chapter 3: Part of Speech Knowledge

## Introduction

Many students have been taught in basic grammar classes the functions of nouns, verbs, adjectives and adverbs. Unfortunately, students may lack the practical knowledge to determine the part of speech of a given word in a sentence, which we know to be an essential skill for vocabulary knowledge. Nation (2001) says, "In order to use a word it is necessary to know what part of speech it is and what grammatical patterns it can fit into. Many linguists now consider the lexicon to play an important, if not central role in grammar" (p. 55). Due in large part to corpus linguistics, we now know that vocabulary learning and grammar are not completely separate skills. Such knowledge has increased our understanding of words.

John Sinclair (1991), the father of corpus linguistics, believed that lexis and grammar are indistinguishable. He also believed that meaning and patterns in language are inseparable. Corpus linguists use the term lexicogrammar to describe the interconnected relationship of grammar and words, and the difficulty of separating the two. There is a difference between the meanings of take in as in "I took in an orphan child," and take off, as in "The plane took off the ground at 2 pm ," the former being figurative and the latter literal. The pattern of the prepositions following the verbs with these words determines the meaning. Also, the reason some verbs are followed by gerunds (e.g., admit lying) and some are followed by infinitives (e.g., asked to sit down) is because the word itself determines the grammatical pattern. Corpus linguistics has the ability to reveal patterns about the language, and specific words, and therefore is ideal for helping language learners understand lexicogrammar rather than memorizing long lists of exceptions to grammatical rules.

By learning a word's part of speech, students can better understand how to identify and look up unknown words, distinguish between different definitions of a word, and learn how to use new vocabulary words correctly. If students are not able to determine the part of speech of an unknown word in a sentence, they will also not be able to understand its collocations, which are sorted by part of speech. Lower-level students are likely to know that sentences have subjects and verbs, even if they do not know the formal names of noun, verb, adjective and adverb. Teachers can build off of this knowledge to help students learn not only the definition for words, but also the parts of speech. Many teachers tell students to learn a word's meaning from context. As Keith Folse (2004) points out, this is much more difficult for non-native speakers, particularly because the context is not very revealing for the meaning of the word. However, surrounding words do give clues about the part of speech of words (see the activity: Identifying Parts of Speech in the Chapter Appendix). Having a part-of-speech knowledge will aid students in finding the correct definition in a dictionary, using the word correctly in speech and writing (Markovic, 2002), and understanding how the word fits into the context grammatically.

There is a reciprocal relationship between corpora and part of speech knowledge. To use corpora, students must have some knowledge of parts of speech, but corpus use can increase that knowledge as well. This is particularly true because COCA is tagged for part of speech information, and students who do not have an awareness of that knowledge may struggle to know how to use corpora. Students who lack this awareness will also struggle between sorting through different forms of words that appear in searches. However, due to the tagging and userfriendly tools of COCA, students should be able to increase their part-of-speech knowledge dramatically as they use the corpus.

## Using the corpus

The easiest way to visually show students different parts of speech in COCA is by choosing the color-coded "KWIC" display radio button at the top of the screen (Figure 6). Similar to the "list" radio button that was previously introduced, this will show the frequency of the word or phrase in the corpus as well as list example sentences. In addition to being able to see KWIC example sentences from the "list" button by clicking on the target word, the KWIC display is also color coded for part of speech. This color-coding system yields great advantages to teaching students who are learning to identify parts of speech as well as learning to sort through multiple meanings of words and phrases.
Display

Figure 6. Key Word in Context radio button.

Figure 7. Key for colors used in KWIC radio button.

Part of speech tag colors and sorting instructions can be found by clicking on the question mark next to the "re-sort" button at the top of the KWIC display. The key is repeated here (Figure 7) for convenience. The "re-sort" button is also useful and allows for users to order the words to the left and right of the node word alphabetically to show patterns. This feature is similar to more traditional concordance line programs. A search using the KWIC feature for the verb phrase "rock out," meaning "to have fun," yields results similar to Figure 8. Rock out is tagged as a
verb and a noun in Figure 8 and is generally correct. Users should be aware that the part of speech system is computer tagged and as a result only $99 \%$ correct. Just as second language users struggle knowing when rock out is a whole phrase or separate words, so does the computer. Occasionally teachers might question the part of speech labels, and in that case should rely on their intuitions as proficient English speakers. For example, in Figure 8 rock is correctly identified as a verb in the first and third lines and correctly identified as a noun in the second line. In the fourth line the corpus misidentified it. Though there may be an occasional error it is so infrequent that it does not outweigh the valuable information provided.

| -hour days, being able to | ock out | (a) little is important. \#" I d |
| :---: | :---: | :---: |
| ive to tear all of the sheet | rock out | and everything 3 Mr-ARROYO |
| $y$ 're going to come in and | rock out | and have fun. But I try to exp |
| an plug it up and just like | rock out | and you can annoy the whole |
| id " Baby Elephant Gunn " | rock out | and Pink Panther Theme " sl |
| on, he shows he also can | rock | as well as any country boy arc |
| the Office actress " could | rock out | 目 3 concert or hit the red carr |
| Tickets, \$ 19.96, for the | ck | by Skid Row on Dec. 29 at the |
| , it has to be reliable and | rock out | enough BTUs to melt potafte |
| us the go-ahead, and we | ock out | for forty-five minutes . Bert sa |
| S : Julie, you 're going to | ck | here CHEN : Oh, my God. P |
| and generally learn how to | rock out | in 3 band. Filmmakers Arne Jı |
| CATIE ARNOLD KAUAI, HI | ROCK OUT | IN HAWAII 6 Dig beneath the |
| re sitting on the seagulls | rock out | in the agoon. First time I've |
| ur beer, while the Chicks | rock out | in the wild-haired " Sin Wagon |
| ence, letting the cameras | rock out | in their own freewheeling danc |

Figure 8. KWIC of rock out.

Another way to teach about part of speech information is by indicating the desired part of speech of individual words in queries. Some words are polysemous and can have more than one part of speech for different definitions, so the corpus allows users to specify part of speech for words. For example, a search for "jump" will yield forms of jump as a noun and as a verb.

However by attaching [ $\mathrm{v}^{*}$ ] to the word jump as in "jump.[ $\mathrm{v}^{*}$ ]" only the results of jump as a verb
will appear. The period between jump and [ $\mathrm{v}^{*}$ ] indicates that $\left[\mathrm{v}^{*}\right]$ describes the part of speech of the word being searched for. This is opposed to a space between items in the search string that indicates new words. If one would rather see what verbs come after the word jump, simply delete the period and insert a space. Therefore, searching for "jump [nn*]" (nn* is the code for all nouns) yields noun phrases such as jump rope, jump cut, jump shot, or jump start. In this example, the part of speech tag is used like a wild card and finds any word that fits the part of speech tag.

Similar to the example above with "jump [ $n n^{*}$ ]", wild cards (symbols that search for any character) may be used in basic searches as well. The "word" query box permits up to eight to fifteen slots or characters (depending on the frequency of the words) and up to ten may be wild cards. An asterisk $\left({ }^{*}\right)$ is the wild card marker allowing for any word or character to fill it. Therefore, phrase searches are possible by using wild card slots. For instance, to find what the three most common words or slots are after the phrase I think that one enters "I think that ***" (Figure 9). Notice that punctuation marks like [,] and ['] count for a slot.

| display | SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| Olist Ochart Okwic Ocompare |  | - | CONTEXT | тот ® $^{\text {a }}$ |
| SEARCH STRING | 1 | $\square$ | I THINK THAT, YOU KNOW | 454 |
| WORD(S) It think that * ** | 2 | $\square$ | I THINK THAT 'S ONE OF | 249 |
| COLLOCATES | 3 | $\square$ | I THINK THAT 'S A VERY | 247 |
| POS LIST | 4 | $\square$ | 1 THINK THAT 'S GOING TO | 244 |
| RANDOM SEARCH RESET | 5 | $\square$ | I THINK THAT A LOT OF | 183 |
| SECTIONS SHOW | 6 | $\square$ | I THINK THAT 'S RIGHT . | 182 |
|  |  |  |  |  |

Figure 9. Wild card search for a phrase.

For more specific wild card searches replace the wildcard * with a specific part of speech such as in the jump example. Tags are used for general parts of speech such as all nouns [nn*] or singular nouns [*nn1*]. Tags do not need to be memorized, but can be selected from the part of
speech list below the search query box (Figure 10). Consequently, one can search for the phrase "drive [pp*] crazy," which will show what pronouns appear in the slot rather than allowing for other parts of speech. The results in Figure 11 show the most frequent pronouns that fit in the phrase, drive (someone) crazy.


Figure 10. Part of speech tag list.

| display | 3 | SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| ©list Ochart ¢wic Ocompare |  |  | ■ | CONTEXT | тот ${ }^{\text {a }}$ |
| SEARCH STRING | 7 | 1 | $\square$ | DRIVE YOU CRAZY | 110 |
| WORD(S) drive [pp*] crazy | a | 2 | $\square$ | drive me crazy | 79 |
| COLLOCATES | $\square$ | 3 | $\square$ | DRIVE HIM CRAZY | 23 |
| POS LIST pron.PERS $\ddagger$ | ] | 4 | $\square$ | drive us crazy | 17 |
| RANDOM SEARCH RESET | $\square$ | 5 | $\square$ | DRIVE THEM CRAZY | 11 |
| sections - show | $?$ | 6 | $\square$ | DRIVE YOURSELF CRAZY | 11 |

Figure 11. Part of speech wildcard phrase search.

For ambiguous vocabulary words that change definitions with part of speech, the corpus can help sort through different word forms by showing the tags on words. In the bottom left hand corner of the screen "click to see options" and then under the drop down list for "group by"
select "none (show POS)" (Figure 12). A word like raise entered into the "word" query box that shows the POS tags looks like Figure 13. The results show how the type raise has different forms and different part of speech tags in the corpus and the frequency for each tag. The most frequent form of raise is the infinitive verb form, followed by the base verb form and singular noun form. This feature of showing POS tags is for the more grammar and corpora savvy. It might be more appropriate for advanced grammar classes learning about the more specific parts of speech.

| SORTING <br> MINIMUM | FREQUENCY - |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
|  | FREQ | LEMMAS WORDS |  |  |
|  | FREQ |  |  |  |
| HIDE OPTIONS |  | NONE (SHOW POS) |  |  |
|  |  |  |  |  |
| \# HITS |  | BOTH WORDS BOTH LEMMAS |  | 10C |
| GROUP BY |  | NONE | (SHOW POS) |  |

Figure 12. How to turn on part of speech tags on words.

| 1 | $\square$ | RAISE (VVI) | 20736 |
| :---: | :---: | :--- | :---: |
| 2 | $\square$ | RAISE (VV0) | 6640 |
| 3 | $\square$ | RAISE (NN1@) | 900 |
| 4 | $\square$ | RAISE (NN1@_VV0) | 738 |
| 5 | $\square$ | RAISE (VVO_NN1@) | 690 |
| 6 | $\square$ | RAISE (NN1) | 7 |

Figure 13. Frequency search showing part of speech tags.

## Applications and sample lessons

Teachers who are unsure how to introduce part of speech knowledge to students might consider teaching students to look for context clues around words and examine part of speech patterns in concordance lines. Students with a strong grammatical background will already be familiar with the role of the parts of speech, but perhaps unsure of how they connect to the words
around them. The "Identifying Parts of Speech Activity" (Chapter Appendix) gives an example of how this might be practiced and taught. Students without a strong grammatical background might benefit from the different viewpoint in instruction.

Part of speech knowledge can also help students who confuse different forms of the same word. For example in Mrs. Brannon's intermediate English for academic purposes class, a student asks which word, racism, or racist should complete the following sentence. "They weren't allowed to vote because the people were $\qquad$ ." In order to help show the difference she searches both words in the corpus on the KWIC radio button and shows her student the results, and asks him to identify what he sees as the differences between the words and their usage. The color-coded part of speech tags show the different parts of speech of the node words as well as how they syntactically fit in with other words. This method also avoids the simple yet incorrect response that "one is a noun, the other is an adjective" because racist may be a noun as well as an adjective. Upon looking at the examples, the student is able to find the answer, while drawing upon his own knowledge about parts of speech and how words are used. This example is simple, and yet shows how learning being informed and driven by data does not need to be a complex process.

Mrs. Wertz' vocabulary quizzes focus on students being able to use part of speech knowledge and meaning from context. For part of the quiz, Mrs. Wertz chooses sentences from COCA that contain the vocabulary word. She makes sure that the context of the sentence does not have vocabulary words more difficult than the one she is testing and that the word is used in a way similar to what they studied in class, modifying them if needed. She gives students a word bank and encourages them to identify the part of speech of the word in the word bank and for the missing word in each sentence. Then the words to choose from are reduced for each sentence by
eliminating the other words in the bank with different parts of speech. Students then choose from the few words that fit the part of speech description and pick the appropriate word based on meaning in context. She has practice exercises similar to the quiz with her students before so that they are aware of the strategies and the format. Her quiz looks something like the following.

| Name: | Vocabulary Test |  |
| :--- | :--- | :--- | :--- |
| 1. anxious | 6. energetic | 11. flavor |
| $2 . ~ s t r e s s e d ~$ | 7. miserable | 12. relaxed |
| $3 . \quad$ calm | 8. seasoned with | 13. indigestion |
| 4. irritable | 9. negative | 14. whole grain |
| $5 . ~ d e l i c i o u s ~$ | 10. upbeat | 15. serving size |

1. A $\qquad$ of breakfast cereal generally ranges from $1 / 2$ a cup to 1 cup.
2. Eat $\qquad$ crackers and an apple before you leave the office.
3. His adult sister and brother were, at this moment, making his life as
$\qquad$ as possible.
4. It is for people who want to be strong, healthy and $\qquad$ .
5. Only then would they taste the pure $\qquad$ of rice and feel its solid goodness in their bellies.
6. Rocco, how did you stay so still and $\qquad$ ? I was nervous just watching you.
7. Sara had left for school feeling edgy and $\qquad$ . As the morning wore on her dark mood deepened into depression.
8. Skiers interviewed by Udry et al. reported that their coaches had a $\qquad$ influence on their recovery.
9. Swimmers must also be careful to avoid foods that cause $\qquad$ .
10. Take a deep breath. When we are $\qquad$ and tense, we take shallow, rapid breaths.
11. The ocean was $\qquad$ , peaceful, the water a brilliant blue.
12. The restaurant also offers a catch of the day and whole fish $\qquad$ herbs.
13. The result is a gentle, fragile soup that is tasty and $\qquad$ .
14. We see more and more teens who are moody, depressed, $\qquad$ and angry because they aren't getting enough sleep, " says Marcu.
15. When you approach life with an $\qquad$ attitude, you set yourself up for greater joy and satisfaction

## Chapter Appendix

## Review

## NOUNS

$\therefore$ Nouns are usually a person, place, thing, or idea.
$\therefore$ In a sentence nouns can be the subject, object, or object of the preposition.
$\therefore$ You can recognize PLURAL nouns because they usually have S or ES on the end (ex. Rose $\rightarrow$ Roses)
$\therefore$ Common endings that nouns have are -ity, -ship, -ion, -dom (ex. Kingdom, charity, friendship)

## Patterns

Nouns can occur in many different places. Here are a few examples, the blank line $\qquad$ represents a noun.
$A d j=$ adjective, $a d v=$ adverb, det=determiner, aux= auxiliary or helping verb, $\mathrm{pp}=$ preposition

| Det + | The (det) + President (noun) |
| :---: | :---: |
| Det + Adj + | The (det) + American (adj) + president (noun) |
| + verb | President Obama (noun) + promised (verb) |
| _ ${ }^{\text {verb+ }}$ | President Obama (noun) + promised (verb) + <br> Americans (noun) |
| PP + | on (pp) + television (noun) |
| PP+ adj+ | on (pp) + public (adj) + television (noun) |

## VERBS

$\therefore$ Verbs are usually an action or a state of being
$\therefore$ All sentences have verbs
$\therefore$ You can recognize different verb endings such as -ed, -ing, -s (ex. Wanted, taking, hurts)

## Patterns

Verbs can occur in many different places. Here are a few examples, the blank line $\qquad$ represents a verb.
$A d j=$ adjective, $a d v=$ adverb, det=determiner, aux= auxiliary or helping verb, $\mathrm{pp}=$ preposition

| Det + Noun $+\ldots$ | The (det) + subway (noun) + stinks (verb) |
| :--- | :--- |
| Noun $+\ldots+$ Noun | He (noun) + smokes (verb) + cigarettes (noun) |
| Aux $+\ldots$ | Is (aux) +chewing (verb) |
| $\ldots+$ Adv | I (noun) talk (verb) quickly |
| $\ldots$ ! (as a command) | Stop! (verb) |
| $\ldots$ to $+\ldots$ | Want (verb) + to spy (verb) |

## Practice Activity

First, identify the part of speech of the words around the highlighted word. Then, identify the part of speech of the word. The highlighted word will always be a noun or a verb.

Adjectives $=$ adj, Verbs $=v$, auxiliary verbs $=$ aux, nouns $=n$, adverbs=adv, determiners $=d e t$, infinitive marker $($ to $)=$ inf, conjunction= conj preposition=prep

1. He was a (der) tough (adj) act ( $\qquad$ ) to (inf) follow (v), but she hoped that someday someone would come along.
2. He didn't look (v) or (conj) act ( $\qquad$ ) like (adv) any (adj) mechanic (n) Id ever seen.
3. Ten minutes ago her mother had been healthy (adj) enough (adv) to (inf) insult (___) and (conj) complain (v).
4. This should have been cause for great relief; to some it was almost (adj) an (der) insult. $\qquad$ _)
5. Kat usually made a splash at (___ ) each (___ ) function (___ ).
6. You make sure hydroponics continue ( $\qquad$ ) to ( $\qquad$ ) function ( $\qquad$
7. If you choose not to frame your (___) beds (__ ), shape (___) the soil so that it is flat-topped.
8. A bouquet of flowers becomes a (___) spherical (___) shape, (___) or an egg shape.
9. We have seen a (___) tenfold (___) increase (___) in (___) the number of men seeking ordination.
10. There are simple things you can do ( $\qquad$ ) to ( $\qquad$ ) increase ( $\qquad$ ) chances $\qquad$ ).
11. I let (___ ) myself (___ ) ) value ( $\qquad$ ) the ( $\qquad$ ) opinions ( $\qquad$ ) of some people more than others.
12. Mayo also provides (____) tremendous (___ ) value (___) for (___) ordinary (___) care.
13. Mike Tyson's daughter was put on (___) life $\qquad$ ) support (___) after ( $\qquad$ ) a ( $\qquad$ ) tragic accident.
14. We have the ability to become a nation of neighbors (___) committed (___) to (___) support (___) and (___) affirm (___) our new leadership
15. This same tape is also used (___) to (___) repair (___) any (___) holes (___) that form
16. Once you lose ( $\qquad$ ) that ( $\qquad$ ) trust ( $\qquad$ ), you can never get it back.
17. For example, some (___) companies (___) request ( $\qquad$ ) an ( $\qquad$ ) e-mail ( $\qquad$ ) address ( $\qquad$ ) to receive information you want.
18. I ( ) can (__ ) state proudly that the entire landscape for a person being diagnosed with MS has changed.

## Chapter 4: Morphological Knowledge

## Introduction

In addition to helping teachers identify frequent words and vocabulary connections to grammar, corpora can also be used to raise students' metalinguistic awareness of morphology. From the Greek roots "morph" for shape or form and "ology" meaning study of, morphology is the branch of linguistics that studies word forms. In terms of metalinguistic awareness this particularly refers to students having an understanding of basic morphology (roots, prefixes, suffixes, etc.), phonology, or the sound system of a language, and phonemic awareness, or how written and spoken forms combine. Nation (2001) identifies four aspects of morphology that are worth teacher instruction. First, students need to be able to recognize word parts and break down words correctly. Second, learners need to recognize what affixes mean and do to words. Third, learners need to be aware of both written and spoken changes when an affix is added. Fourth, learners need to know which morphemes can be combined and which cannot, or which roots can combine with which affixes (p. 275-278). Koda (2000) adds that good readers have strong metalinguistic awareness because they understand that phonological and orthographic forms are connected, words can be divided into smaller, meaningful parts, and that meaning can be obtained from smaller parts. Therefore, good readers know how to break down words into parts to find meaning. Ferris (2009) supports this idea and says that students who need support with academic reading should be taught English morphology to analyze the parts of a word for meaning. Additional research has demonstrated that knowing derivational morphemes increases reading ability and vocabulary growth (Markovic, 2002).

Morphology is traditionally divided into two categories: derivational and inflectional. Derivational morphology is how a word breaks down into its base or "root" form and any affixes
attached to it, i.e. prefixes or suffixes. Derivational morphemes can either give the root meaning of a word, such as script in scripture, which means "write," or the part of speech, such as the noun ending, ment, as in pavement. In addition, some derivational morphemes are used to convert words of one part of speech to another, such as the suffix ize, which makes nouns become verbs, as in idolize. The second category, inflectional morphology, marks the grammatical endings on words. In English these are things like plural "s", "ing", third person singular "s", and past participles "ed" and "en." They do not change the part of speech of a word, but inflect a word for grammatical functions. Teachers, as well as their students, may not be as familiar with morphology as frequency and parts of speech. Therefore, two full-length lesson plans (Derivational Morphemes Part I and Part II) in the Chapter Appendix are included for precorpus instruction to introduce these ideas.

Though some students may be aware of English endings such as "ed" and "en" for specific grammar rules (inflectional) or "un" or "dis" for some vocabulary words (derivational), that is often the end of their morphological knowledge. Morphological knowledge can help students learn how the parts of words have meaning on their own, but combine together to create new meaning for the whole word. The meaning of these parts can then be used as students find their own simplified definitions for words. For example, students may know what play and full mean individually. Combined together, playful does not mean someone is full of play, but rather than he or she is fond of amusement and fun. The student can then use the word parts to make his or her own definition that playful means you like to play (Nation, 2001).

Morphological knowledge can also help students see relationships between related word forms, both grammatically and semantically. Grammatically, a lemma is a single base word and all of its grammatical inflections. Understanding that decide, decides, decided, and deciding are
all related grammatically is not obvious to all students. However, a lemmatized search in COCA yields all grammatically related forms of a word and may aid students in making those connections. This can especially be useful in the difficult task of inflecting phrasal verbs and idiomatic expressions. In terms of semantic relationships, Nation (2001) recommends that advanced students study word roots to learn new words, particularly ones that correspond to their specific area of study. Thereby, students may learn that lucid and lucent both related to giving off light because they are based on a common root. Additionally, language learners can break down words into segments and then use their background knowledge of other words and their meanings to relate them to new words and morphemes. Consider the word transient. Trans is related to translate, transfer, transit, all which have the idea of movement. Ent is related to different, recent, current all of which are adjectives. A student who has morphological knowledge can then guess that transient is an adjective related to movement. By relating new words to previously known words, and breaking words into parts, students can make educated guesses about meaning, and increase their ability to remember the meaning by linking the knowledge to things they already know. Moreover, morphological knowledge can also help students differentiate between commonly confused words. Though part of speech knowledge and context can help students know the difference between success and succeed and successful, morphemes show that distinction as well, and provide another way to remember. Therefore, morphemic knowledge provides several effective vocabulary learning strategies for students. Using DDL to learn morphemes and expand morphological knowledge can be a great way to practice these strategies.

Like part of speech knowledge, it is helpful to have some basic understanding of morphology to use a corpus, and using the corpus may also help develop this knowledge.

Knowing how to segment words into parts is particularly helpful because students will be able to better narrow corpus searches. If a student is unsure whether he wants to say likely or likeable a search of "like*" in COCA will yield all of the possible words that have like at the beginning. This is only possible though if students know how to find the root and break off the morphemes $l y$ and able.

## Using the corpus

Wildcard searches can help students recognize morphemes as they look at examples or search for related words in a corpus. For example, someone using the corpus to guess the meaning of teleportation would first have to break down the word into the morphemes tele, port and ation. Then, each morpheme could be searched for in the corpus to find a related word that might hint at the meaning of the part, and aid in creating meaning for the word as a whole. This can be done with a wild card search in COCA to find related words to "tele*" *port" or "*ation" in the corpus. The user can then select words like telephone, airport, or transportation in trying to find related words. These wildcard searches are especially useful for derivational morphemes.

Part of speech (POS) tags in COCA can aid teachers in the instruction and finding of inflectional morphemes. Tags range from the general categories of noun [nn*], verb [ $\mathrm{v}^{*}$ ], adjective [ $\mathrm{j}^{*}$ ], and adverb [ $\mathrm{r}^{*}$ ] to very specific categories such as plural nouns [*nn2*], third person singular verbs [v?2*], past tense verbs [v?d*], or past participles [v?n*]. Tags do not need to be memorized but can easily be selected from the drop down menu POS list. By clicking the question mark to the right of the POS list, a help screen appears that contains a key to all of the part of speech tags used in the corpus. Therefore at even a beginning level teachers can help to develop their learners' morphological awareness by searching for a single tag such as ing words " $\left[\mathrm{v} ? \mathrm{~g}^{*}\right] "$ and show the resulting words to students to find a pattern (Figure 14).

| DISPLAY | $\square$ | SEE CONTEX |  | RDS + [CONTEXT] |
| :---: | :---: | :---: | :---: | :---: |
| ©list 〇chart KWIC ©compare |  |  | - | CONTEXT |
| SEARCH STRING | $\square$ | 1 | $\square$ | GOING |
| WORD(S) [v? ${ }^{\text {[2*] }}$ | $\square$ | 2 | $\square$ | BEING |
| Collocates | $\square$ | 3 | $\square$ | DOING |
| POS LIST verb.ING $\dagger$ | $\square$ | 4 | $\square$ | TRYING |
| RANDOM SEARCH RESET | $\square$ | 5 | $\square$ | LOOKING |
| SECTIONS $\square$ show | $?$ | 6 | $\square$ | HAVING |

Figure 14. Wild card search for grammatical forms.

In order to search for the inflectional morphemes of a word, perform a lemmatized search by placing a base form of a word in brackets. The lemma will yield all the inflected forms and their frequencies. The following results show all the forms of study by placing "[study]" into the word search box (Figure 15). It results in study, studies, studied and studying, with the frequencies of each form given. Lemmatized searches can be combined with frequency searches in order to look for all of the possible forms of a word in a phrase. To find out if it is more common to use a phrasal form run errands in simple past, simple present, or present progressive simply perform a lemmatized search "[run] errands" and look at the frequency of all forms (Figure 16).


Figure 15. Lemmatized search of study.

| DISPLAY | 3 | SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 〇list Ochart Okwic Ocompare |  |  | - | CONTEX | тот $\quad$ - |
| SEARCH STRING | 3 | 1 | $\square$ | RUN ERRANDS | 122 |
| WORD(S) [run] errands | - | 2 | $\square$ | RUNNING ERRANDS | 107 |
| COLOCATES | ¢ | 3 | $\square$ | ran errands | 32 |
| POS LIST | 3 | 4 | $\square$ | RUNS ERRANDS | 14 |
| RANDOM SEARCH RESET | $\square$ |  |  | TOTAL | 275 |
| Sections - show | 3 |  |  |  |  |

Figure 16. Lemmatized phrase search of run errands.

## Applications and sample lessons

Students at an intermediate/advanced level are typically familiar with inflectional morphemes. Though they may not know them by that name, they have typically used them in their grammar classes. Students who have a language background in the Romance languages might be able to see root word connections to their own language and understand derivational morphemes more quickly. Therefore, teaching these concepts is not always complicated and in some ways can help students remember words more quickly. Consider the following case study of a teacher helping students use morphemes to relate vocabulary words to their own lexicon.

Ms. Read has just been asked to review a list of ten words with her academic English writing class. She chooses the first five to review in class: previous, resident, perspective, conference, and motivation. She wants her class to be able to relate these words to words they already know to help them remember them. After briefly reviewing definitions, she draws Table 2 on the chalkboard. She reserves a column for the word, its parts, and related words.

Table 2

Breaking Down Vocabulary Words into Parts (Incomplete)

| Vocabulary Word | Parts of the Word | Related Words |
| :--- | :--- | :--- |
| previous |  |  |
| resident |  |  |
| perspective |  |  |
| conference |  |  |
| motivation |  |  |

Ms. Read then invites the whole class to decide how the words might be divided into different parts. She reminds her students that they are not dividing words into syllables but where they think the root of the word is. She writes the students' suggestions in column two, not correcting or ignoring them. That way, the results are not limited and the students can learn how to divide words from the results of their searches and discover for themselves. Also, while searching in COCA, it is sometimes helpful to have varied searches available, in case the results are not as expected. Table 3 shows her students' suggestions.

Table 3

Breaking Down Vocabulary Words into Parts (Partially complete)

| Vocabulary Word | Parts of the Word | Related Words |
| :--- | :--- | :--- |
| previous | previ+ous, pre+vious |  |
| resident | resid+ent, res+ident, re+sident |  |
| perspective | per+spec+tive, perspect+ive |  |
| per+spect+ive |  |  |
| conference | con+ference, confer+ence |  |
| motivation | motivat+ion, mo+tivat+ion |  |

Finally, Ms. Read uses wild card searches in COCA to search for related words using the breakdown her students have created. She searches for "previ*" as well as "*ous", "pre*" and "vious." The class selects two words, preview, and obvious that they feel are related based on the results; Ms. Read merely guides them in the right direction. As a class, they are able to fill in the table from the various searches and they discuss how the words might be similar or related to remember the new words, noting that there are sometimes slight spelling changes. Ms. Read does not seem to mind that the students may not have correctly identified the root of the word in the second column as long as they do so in the third column with the related word. Note that the "Related Words'" column has words related to the different morphemes, not just the root morpheme. The final table looks like Table 4.

Table 4

Breaking Down Vocabulary Words into Parts (Complete)

| Vocabulary Word | Parts of the Word | Related Words |
| :--- | :--- | :--- |
| previous | previ+ous, pre+vious | preview, obvious |
| resident | resid+ent, re+sid+ent, re+sident | reside, president |
| perspective | per+spec+tive, perspect+ive, <br> per+spect+ive | respect, relative |
| conference | con+ference, confer+ence | confess, experience |
| motivation | Motivat+ion, mo+tivat+ion | motor |

The results in Table 4 reflect the knowledge of the students in Ms. Read's class, and might be different, or incomplete, depending on the students. She helps facilitate the students' learning and creativity as well as reminds them of things they already know when they seem to get stuck. For weekly homework, Ms. Read has her students complete the same activity for the other five words.

Other classroom applications may help students see the relationship between words. Words that are spelled similarly are often confused such as the example previously given of racism and racist. Morphemes can help students learn which word is which (-ism is always used for nouns), but some words, that seem to have the same root because of spelling similarities, may not be related, or maybe in the past were but the meaning has changed. Neither teachers, nor students, have to be experts or etymologists for each word, just as long as they are aware of the possible issues. DDL can help generate activities for the teacher to introduce the idea of related words to students. The teacher's responsibility is then to introduce aspects of vocabulary words
in the activities, generate discussion and thinking (to build metalinguistic awareness) in class, and let students study and explore them more on their own. Consider the following warm-up vocabulary exercise to be done in class (the correct answer is marked with a *):

1. Circle the word that is not related in meaning. clarify clarinet clarity clark*
2. Circle the word that is not related in meaning annotate connote denote keynote*
3. Circle the word that is not related in meaning prospect circumspect* inspect suspect

All of these activities were created by a corpus wildcard search. For the first example the search string was "clar*"; for the second example the search string was "*note"; for the third example the search string was "*spect." The teacher must simply identify the root of the word, while remembering that there are sometimes spelling changes such as with clarify (the root is clarus, or clear), and then search for it. All of the words in example 1,2 and 3 seem to be related by spelling but are not necessarily related by meaning. Though in example 3 all of the words come from the same Latin root, the meaning of the words is not related for current English usage. Intuitions as a native or near-native English speaker are still needed to find words linked by meaning and to lead a class discussion.

## Chapter Appendix:

Lesson Plan: Derivational Morphemes Part I

Goal: Help students understand how they can identify a word's part of speech through derivational morphemes and how that knowledge can be useful.

## Learning Outcomes:

- Students will be able to identify common derivational morphemes.
- Students will know how to match derivatives with related words in a dictionary
- Students will demonstrate producing words from different derivational endings


## Materials:

Derivatives handout
Word parts cards
Dictionaries (pref. English only dictionaries)
Dictionary worksheet
Class vocabulary lists

## Procedures:

1. (5 min) Introduction: Introduce the idea of parts of words by handing out to different students parts of a word from the word parts cards. Have them try to piece the words together. Once the class has the two parts of the word together, ask them to categorize each word into its part of speech.
2. (20 min) Instruction: Go over the Derivatives handout as a class. You might want to have them work in pairs or groups to separate the roots and endings to evaluate understanding. Give them the list of derivational endings, stressing that they don't need to memorize them, just be able to identify them.
3. (15 min) Receptive practice: Using a class vocabulary list or words they are studying have students look up several of their vocabulary words and identify the different derivations and endings for those derivations that correspond to parts of speech. To facilitate this you might have them fill out Dictionary worksheet and do one example together as a class. Afterwards as a class map some word webs on the board of different word derivations students found.
4. (20 min) Productive practice: Divide students into teams. Give each team 2 minutes to review the derivational endings and then put them away. The object of the activity is for each team to be the first to identify the part of speech of each derivational ending given and an example word. Students may not use examples given on the handout, but you may allow them to use a list of vocabulary such as the Academic Vocabulary List or a dictionary. Write an ending on the board and the first team to raise their hand with the correct part of speech and an example receives a point.
5. (15 min) Contingency Activity: Some derivational morphemes are more productive than others. You may want to demonstrate to your class how $C O C A$ can find a list of example with morphemes by wildcard searching *able, or *(your favorite affix here). As a class explore which morphemes are most productive for each part of speech.

## Derivational Morphemes

In English many words have a root, or most basic form of a word, shared with other words. If you know one word and its part of speech, it is easier to know the other forms of that word by adding endings to the root. That makes it easier to learn many words at a time instead of just one word at a time.

For example:
Adjust is a verb that means to make a small change to make something better. Adjust is a root.

If you already know that -ment is an ending used for nouns, you can add that ending (adjust-ment) and learn that adjustment is the noun form.

If you already know that -able is an ending for adjectives you can add that ending (adjust-able) and learn that adjustable is the adjective form.

If you already know that -er is an ending for nouns that means someone who does an action you can add that ending (adjust-er) and learn that adjuster is another noun form.

Therefore by learning derivations and derivational ending of words you learned four words instead of one. Derivational endings above are -ment, -able and -er.

In the following examples try to separate the roots from the derivational endings. Some words can have more than one derivational ending, in this exercise you only need to identify one.
Word
Root
Ending

1. Valuable
2. Responsibility
3. Creative
4. Imitation

Value
able
Responsible
Create
5. Offender
6. Violence

Violent
7. Professional
8. Powerful
9. Generation

Generate
10. Representative

Each part of speech has different derivational endings. The following are common derivational endings for each part of speech. By being familiar with these endings you can identify words with similar roots and learn more words at once. NOTE: Some derivational endings overlap. For example: y is an ending for both adjectives and nouns.

## Adjectives

Derivational Ending
al/ial
ious/ous
ture/ure
an/ian
ar
ate
ful
ic
y
il/ile
ary
ish
ive
able/ible/ble

Example
mental, partial
numerous, serious
future, unsure
American, historian
similar, lunar
corporate, immediate
painful, bountiful
acidic, ethnic
plenty, happy
tranquil, juvenile
primary, necessary
British, selfish
native, positive
likeable, possible

Adverbs
ly
ward
wise
probably, really
backward, heavenward
otherwise, clockwise

## Verbs

ate
en
ify/fy
ize/ise

Nouns
ity
ist
let
ment
sion/tion/cion/ation/ion
age
al
ance/ence
ant
mony
ory
sis
y
ure
er
hood
participate, communicate
flatten, harden
justify, modify
advertise, patronize
majority, superiority
journalist, capitalist
booklet, wallet
statement, acknowledgement
education, suspicion
bandage, adage
animal, tribunal
substance, permanence
abundant, aberrant
ceremony, alimony
auditory, crematory
sepsis, analysis
country, party
treasure, culture
baker, teacher
parenthood, priesthood

Word parts- introduction


## Dictionary Worksheet

Using the vocabulary words provided by your teacher consult a dictionary to fill out the following sheet. When looking in the dictionary look for derived forms, derivatives, or the words before or below the dictionary entry to find alternate forms.

| Word | Part of <br> Speech | Alternate Forms | Derivational <br> Endings | Part of Speech |
| :---: | :---: | :---: | :---: | :---: |
| Example: <br> Probability | Noun | Probable, probabilistic, probabilism | able, ic, ism | Adj, adj, noun |
| 1. |  |  |  |  |
| 2. |  |  |  |  |
| 3. |  |  |  |  |
| 4. |  |  |  |  |
| 5. |  |  |  |  |
| 6. |  |  |  |  |
| 7. |  |  |  |  |
| 8. |  |  |  |  |
| 9. |  |  |  |  |
| 10. |  |  |  |  |

## Lesson Plan: Derivational Morphemes Part II (Greek and Latin Roots)

Goal: Help students understand how they can guess the meaning of a word by knowing common Greek and Latin roots.

## Learning Outcomes:

- Students will be able to identify several Greek and Latin roots.
- Students will learn how to find Greek and Latin roots
- Students will classify words with graphic organizers


## Materials:

Discovering Greek and Latin Roots
Some Common Greek and Latin Roots
Dictionaries
Word Families
Sample word map

## Procedures:

1. (10 min) Introduction: Write on the board the following groups of words. Have students guess which word in each group doesn't belong, and have a reason for their guess. You might want to have them write the groups down and give them a few minutes on their own, and then after a few minutes let them use their dictionaries to guess from definitions.

Group 1: concise, decide, imprecise, incise, accident
Group 2: suggest, aggressive, gradual, progress, grade
Ask a few students to volunteer answers as to what they think the answer is. Tell them the answer and that the lesson today will be an easy way to know which words are related and which one isn't.
2. (20 min) Instruction: Give students a copy of the Greek and Latin Roots handout. Go over the answers to the introduction with them and then familiarize them with some of the more common Greek and Latin Roots. Show them how to graphically organize the related words so that by learning a root you can show the relation to other words and learn more words.
3. (20 min) Receptive practice: Teach students one or a few sources (depending on their level) where to find Greek and Latin roots about a word so that they can learn related words. Have them practice looking up vocabulary words and filling out the handout "Word families"

Sources to learn about related words:
A. http://wordinfo.info/-

Type the vocabulary word into the cross-reference search area.

The results will show a few options for definitions and then at the end the etymology. Then look for the roots under the phrase "The word, or words, you searched for may be seen in the following family unit, or family units." If you click on those suggestions it will give you a list of related words in the same family from the same root.

## B. Oxford English Dictionary Online

Type the vocabulary word into the search field. At the top of the entry choose etymology to find the root. If the root is a hyperlink it will lead you to related words, or the dictionary entry of the root to give you other examples. Or if you only know the roots, look at the wiktionary page to find related words.
C. Wiktionary: English words by Latin roots

This site is in alphabetical order by the Latin root. If you already know the Latin root you can see examples of related words. Because this is an openly edited site, you may want to check a dictionary to make sure they really do have similar roots and definitions.
4. (10 min) Productive practice: Have students select a few of the words from the "word families" sheet. Have them practice mapping associated words with derivational endings and from similar roots. You can have them draw their own graphic organizer on blank paper. Make sure they write short definitions for related root words when they draw them on or below the graphic organizer.

## 5. (15 min) Contingency Activity:

Give students words from a vocabulary list that you are working on or choose some example words from "Some Common Greek and Latin Roots" handout. Divide students into small groups and choose a 3 of those words to write on the board. Have students (using dictionaries and other tools) find as many words with similar roots (any of the roots). The group with the most unique words (words that other groups don't have) wins the round. Repeat with words in groups of 3 .

## Discovering Greek and Latin roots

Many English words are related to Greek and Latin words. The root or basic meaning of many English words comes from Greek and Latin words. Those who know a Latin-based language like Spanish, French, Portuguese, Italian or Romanian might recognize some English words to be similar to words in their own languages. Words that are similar between languages are called cognates.

Even if you do not know a Latin-based language, learning some of the most common Greek and Latin roots can help you learn English. By combining the roots with the derivations learned earlier you can guess the meaning of unfamiliar words and learn many words at once by mapping them together.

## For example:

Concise, decide, imprecise, and incise are all related and come from the Latin word caedere that in English became cis or cid meaning to cut.

Concise means to cut out detail and give a short statement.
Decide means to make a final choice, or cut off hesitation about something
Imprecise means not exact, or not cut correctly
Incise means to cut into something, or carve

By learning just one root cis/cid you learned four English words.
Now that you know the answer to group 1, can you guess the answer to group 2?

Aggressive, gradual, progress, grade all come from the Latin root gradi meaning to step or go. In English it changed into many different forms like gress and grad.

Aggressive is something or someone who is ready to attack or go forward with action
Gradual is moving step-by-step, or moving in small amounts
Progress is to go forward
Grade is a position in rankings, or one step above another.

Using a dictionary can you guess the root of these words? Colloquial, eloquent, loquacious. What do they all mean? By learning words in groups that are similar in meaning, you do not have to memorize each word but instead you can make a drawing to remember how all the words are connected. Here is an example of a drawing:


By using the knowledge about derivational endings and words from the same root you can add more related words and connect them also. Here is an example of a drawing with related root words and derivational endings:


Learning new words different parts and how other

Regression you where to find similar words from the same root.

## Some Common Greek and Latin Roots

Sometimes the Greek and Latin words look different in English from their original form. The sample English forms are examples of how the word changed when it was borrowed into English. There might be other forms than what are listed. This list is an example of root words. There are many more to learn.

| Origin | Root | Sample English forms | Meaning | Examples |
| :---: | :---: | :---: | :---: | :---: |
| Latin | Mors | Mors, mort | Death | Immortal, mortify |
| Latin | Audio | Aud, audi, audio | To hear | Audience, audition |
| Latin | Bene | Bene, bon, beni | Good, well | Benefit, benign |
| Greek | Biblio | Bibl, biblio | Book | Bible, bibliography |
| Greek | Bio | Bio, bial, bian | Life | Biography, biopsy |
| Latin | Caput | Capt, capit, cap, chap | Head | Capital, chapter |
| Latin | Dominus | Domin, dom, dang | Master, rule | Dominate, danger |
| Latin | Frangere | Frag, frac, frail | To break | Fragile, infraction |
| Latin | Habere | Hab, hibit | To have, hold | Habit, inhibit |
| Greek | Homo | Homo, homeo, | Same | Homonym, |
| Latin | Iuris | Jus, juris, jur | Law, right | Jury, justify |
| Greek | Logy | Ology, logy, | Study, science of | Biology, chronology |
| Latin | Lucere | Lustr, luce, luc | To shine | Elucidate, lucent |
| Greek | Mant | Mant, manc | Prophecy | Necromancy, mantic |
| Latin | Mittere | Mit, miss, mitt | To send | Transmit, omit |
| Latin | movere | Mov, mot, mob | To move | Mobile, remove |
| Latin | Noscere | Cogn, conn, gnor | To know | Cognition, ignorant |
| Greek | Onym | Onom, nym, | Name | Synonym, antonym |
| Latin | Pellere | Pel, peal, push, pulse | To drive, push | Propel, repulsive |
| Latin | Pendere | Pend, pens, pond | To hang, weigh | Suspense, ponder |

## Word Families

Using the vocabulary words provided by your teacher find the roots and related words of your vocabulary words. Fill in the work sheet. Try and find several related words that will be useful to you.

| Word | Part of <br> Speech | Root (s) | Meaning of Root(s) | Related words |
| :---: | :---: | :---: | :---: | :---: |
| Example: <br> Subsequent | Noun | Sub, sequ | Sub-below <br> Sequ- following | Sue, suit, sequestrate, consequence |
| 1. |  |  |  |  |
| 2. |  |  |  |  |
| 3. |  |  |  |  |
| 4. |  |  |  |  |
| 5. |  |  |  |  |
| 6. |  |  |  |  |
| 7. |  |  |  |  |
| 8. |  |  |  |  |
| 9. |  |  |  |  |
| 10. |  |  |  |  |

## Chapter 5: Synonym Knowledge

## Introduction

When defining words, there are several possible approaches that one can take. A word can be defined in terms of what it is not, or by giving its antonyms; happy is not sad. A word can be defined in terms of what categories it fits into, or by its hyponyms; necklace is a kind of jewelry. Likewise, a word can be defined in terms of words that share similar features, or its synonyms; myth is a synonym for legend. However, with any definition, we are constrained to define a word in terms of other words. Each word represents an idea or a concept, and to better understand it, we interpret how it fits in with words and concepts familiar to us (Schmitt, 2000). Synonyms are one way to do this.

Language learners who already understand a concept in their native language can quickly and effectively learn the translation of the word and master it in both languages (Folse, 2004). However, problems arise when students rely on translations too much and fail to use knowledge they already have in the foreign language. Some potential pitfalls could include if the concept does not exist in the native language, if the concept is understood differently within the culture, or if there are multiple words in one of the languages, such as the differences between querer and amar in Spanish, which is only equivalent to love in English. Additionally, if words are only learned in a one-to-one translation method, it limits students' abilities to produce language in lexical variety. Lexical variety is a problem for many learners, particularly in second language writing (Ferris, 2009).

As a proposed solution to such issues, teachers can guide students in developing their knowledge of synonyms. Synonymy is perhaps one of the better-known skills among language teachers, and is related to other aspects of metalinguistic awareness such as specific uses and
semantic preference. Two synonyms can have very similar meanings, but distinct uses depending on context, such as pass away and expire. Knowing similar words in meaning can help develop other areas of metalinguistic knowledge such as register, part of speech, and collocational knowledge that could potentially transfer to synonyms. This is particularly relevant because collocational and register differences are typically what set synonyms apart (Liu, 2010). Also, because two words are hardly ever identical in meaning, synonyms are "...essential for expressing shades of meaning to help us convey our ideas and feelings precisely for effective communication" (Liu, 2010, p. 57). Therefore, language learners who use more synonyms can be more effective in communication and thus avoid misinterpretations of meaning.

While learning synonyms is an important part of knowing a word, and a great learning strategy, there are a few issues that teachers should be aware of. For example, in a mixed background ESL class, a teacher might give definitions in simple synonyms that learners already know, such as remembering that glad means happy. However, if students are unfamiliar with the concept of happy the strategy becomes extremely ineffective. The concept and core meaning must be understood before students can learn any synonyms or distinctions of meanings. Another potential problem with learning synonyms is what Ferris (2009) calls, "thesaurus errors" (p. 107). Students who attempt to use synonyms might err on choosing a synonym for the wrong context, or failing to adjust grammar to the new word. Fortunately, these errors can be corrected with increased student awareness of register knowledge and collocational knowledge, both of which are essential in learning synonyms.

Another consideration with synonyms in vocabulary learning is the possible confusion between their distinct, nuanced meanings, because of the polysemous nature of words. It can be a great disadvantage to students if they assume that synonyms can be used completely
interchangeably. In his study of five near-synonyms, Liu (2010) states, "While synonyms express basically the same concept, they often do so in different fashions...In other words, synonyms are often not entirely identical in meaning and hence not completely interchangeable" (p. 56). Essentially, teachers need to help students learn to distinguish synonyms for words that have multiple meanings, particularly for students of higher skill levels. Like the other "thesaurus errors" that Ferris (2009) mentions, different meanings of a word can be distinguished by specific set phrases, registers, or collocations. By using DDL and examples from a corpus, teachers can help students visually see the differences. Otherwise, students might not understand why you cannot say parade symptoms, but have to say show symptoms if show and parade are synonyms. For this reason, DDL can provide more information than by just using a thesaurus.

## Using the corpus

Synonym searches are easy to perform in COCA because of its built-in thesaurus. To search for synonyms, choose "list" on the radio button at the top of the screen and enter the word into the "word" query box with the formula [=word]. For example, to see synonyms of the word beautiful enter "[=beautiful]" into the search box. The results will show a list of synonyms and their frequencies on the right (Figure 17). Be aware that the corpus is not semantically tagged and does not distinguish between different meanings of a word. Thus, it will show synonyms for all the word forms spelled the same way. For example, the synonyms of bank include several definitions such as financial institution, side of a river, and group of items, among others. Synonym searches in COCA allow learners to practice distinguishing between the multiple meanings of words. Teachers should guide their students through synonym searches and help them to avoid any potential confusion, as well as to know that not all words in the corpus have synonyms (due to the thesaurus used, and exclusion of phrases).


Figure 17. Simple synonym search of beautiful.

Synonym string searches are a unique feature of COCA that allow for other learning possibilities. Synonym searches show the relevant synonyms, in addition to synonyms for those results by clicking on the [s] next to the resulting word (Figure 17) For instance, by clicking on the [s] next to wonderful (in the list of synonyms for beautiful) the corpus will search for synonyms of wonderful. The synonyms strings can be repeated as many times as needed. These searches can be useful to find related synonyms of a word, to create graphic organizers like word webs for vocabulary words, and to help students find simple synonyms for words with which they are unfamiliar.

Another feature of COCA is lemmatized synonym searches. As previously described, lemmas show all of the inflectional forms of a word. In this case it will show all the inflectional forms of synonyms for the desired search. This can be used to see which grammatical forms of synonyms are most commonly used, and thereby educate learners on the use and meaning of words. As long as default settings remain intact, lemmas will be listed in order of frequency. Lemmatized synonyms for inquire (Figure 18) show that the most common synonym is asked followed by the present tense question. To perform a lemmatized synonym search simply add another pair of brackets around the synonym. Instead of searching for [=inquire], enter [[=inquire]] into the "word" query box to see synonyms of all inflected forms.

| ASKED [S] | 152431 |
| :--- | :---: |
| QUESTION [S] | 132509 |
| ASK [S] | 95421 |
| QUESTIONS [S] | 77909 |
| ASKING [S] | 31934 |
| ASKS [S] | 17818 |

Figure 18. Lemmatized synonym search for inquire.

Synonym searches can also be useful in identifying acceptable substitutes in some phrases (for more detail on phrases see Chapter 6). Set phrases and idioms are often tight and invariable, whereas loose, open phrases accept substitutes. Tight phrases include tickled pink, slap happy, or hands down, where none of the words in the phrase are substitutable by synonyms. Loose phrases are dumber than an ox and down in the dumps, both of which can allow for substitutes like dumber than a dog and down in the pits. For example, if students are looking for another way to say understand the problem, they might search for "[[=understand]].[v*] the problem" (Figure 19). The results show the frequencies of the synonyms for understand that occur in similar phrases. An alternate search that would allow more variation and show similar phrases would be "[[=understand]].[v*] the [=problem]". This yields alternative results such as see the difficulty and understand the question.

| DISPLAY | 7 |
| :---: | :---: |
| 〇list ○chart ○kwic Ocompare |  |
| SEARCH STRING | 7 |
| WORD(S) $[$ [ $=$ Inderstand $]$ ].[ $\left.\mathrm{v}^{*}\right]$ the problem | ] |
| COllocates |  |
| POS LIST |  |
| RANDOM SEARCH RESET | $\square$ |
| SECHIONS SHOW | $?$ |
| \|l| |  |


|  | - | CONTEXT | тот |
| :---: | :---: | :---: | :---: |
| 1 | $\square$ | SEE THE PROBLEM | 124 |
| 2 | $\square$ | UNDERSTAND THE PROBLEM | 79 |
| 3 | $\square$ | RECOGNIZE THE PROBLEM | 38 |
| 4 | $\square$ | SAW THE PROBLEM | 33 |
| 5 | $\square$ | RECOGNIZED THE PROBLEM | 30 |
| 6 | $\square$ | IDENTIFY THE PROBLEM | 30 |

Figure 19. Synonym phrase search.

## Applications and sample lessons

COCA is not semantically tagged to differentiate between the synonyms for different shades of meaning of a word; however, neither are thesauri. Native speakers who know the shades of meaning of a word are able to find a synonym that matches the meaning they want. Non-native speakers often face this difficulty, and language teachers often struggle with teaching them how to find the correct synonym, no matter what tool is used. Other aspects of vocabulary knowledge, such as register and collocations, can help resolve this issue, especially through the features in COCA. It can be advantageous to teach synonyms through a corpus rather than standard thesauri because of the ability to choose the most appropriate word based on example sentences. Consider the following example of data-driven synonym learning:

Ms. Perry's low-advanced academic writing class is learning to improve their writing by using more academic words. As a class, she leads them through the following exercise, demonstrating how to find a synonym. Having located the synonym, students then view sample sentences for that synonym to ensure that it is the correct choice. Here is an example of a worksheet she gives her students.

All of the bolded words in the following paragraph need to be replaced by a synonym. Using COCA, choose the best synonym for the word that matches the meaning as used in the paragraph and fill out the attached worksheet (Table 5).

Technology addictions make a person lose face-to-face contact. Sometimes it is hard for an addicted person to talk with society by himself because of the problems that the bad habit causes. For example, people who are addicted to the Internet get angry easily when the network is working slowly or it is interrupted.

These causes might have effects when the addicted person is behaving in society.
Then, it is very important for an addict to count on support from parents, teachers and friends in order to return to his or her normal social life.

## Table 5

Synonyms and Examples of Academic Vocabulary

| Word | Synonym (guess) | Synonym from COCA | Example sentence: |
| :--- | :--- | :--- | :--- |
| make |  |  |  |
| hard |  |  |  |
| problems |  |  |  |
| habit |  |  |  |
| addicted |  |  |  |
| get angry |  |  |  |
| behaving |  |  |  |
| count on |  |  |  |

Ms. Perry begins by asking the students for ideas of synonyms to encourage students to use their own knowledge and intuitions. She then moves to the corpus--shown on a projector connected to a computer--and demonstrates how students can find a good synonym for each of the words in the exercise. She enters "[=make]" into the corpus and a variety of words appear in the results such as be, do, get, kind, become, name, force and many others. Because make is such a common word in English she shows the class how it is possible to narrow the results and search for synonyms of make in this context. Though this is not always necessary, in this case she
decides to enter "[=make] someone" into the search field and results such as be someone, get someone, nominate someone, appoint someone, force someone, elect someone and require someone appear. Based on student intuitions they choose which one to look at first. The class chooses become someone and Ms. Perry goes back to the first results page and clicks on the [s] next to the word become to make sure that the synonyms match for the sentence in the paragraph. The class discusses the possible matches and decides that they don't think become is a good synonym in this sentence. Ms. Perry returns back to the results for "[=make] someone" and the class choses force someone. Once again they perform a synonym string search for force, but this time the results make more sense. Then the class clicks on the word force someone to make sure that the example sentences show a definition similar to the written paragraph. They choose one example sentence where force matches the meaning and fill in the chart. After doing the first word together as a class, Ms. Perry pairs students to finish the activity. After sufficient time, they review the findings as a class so that students feel comfortable in their ability to find synonyms. After the activity, Ms. Perry then gives students selections from their own papers where she has underlined words or phrases that need to be replaced.

Reading teachers choosing vocabulary words from a text might use corpus synonyms to assess vocabulary understanding from a reading passage. Because the corpus is not semantically tagged, a synonym search for a polysemous word will give options for distractor answers. This kind of activity is appropriate after students have already learned the concept of the word. A sample question might look like:

1. Which word is closest in meaning to how indicate is used in the passage?
a. imply
b. reveal
c. signal
d. point

In this case, teachers can use the corpus to generate questions and discussions to help the class understand the meaning of a word in context. Questions such as these are common on standardized English exams and therefore can be beneficial to practice in class. After these questions, a class might explore example sentences or perform synonym string searches to further their study.

Graphic organizers are a common way to organize vocabulary words that are closely related. Synonym searches and synonym string searches can be organized into word maps, or word webs, to show differences and similarities between words. This is especially effective if students are familiar with the different synonyms but not with the distinctions between them. The visual thesaurus at http://www.visualthesaurus.com/ is a good example of word maps that show polysemous synonyms for words. Teachers can create similar maps in their classroom. This is done by putting the target word into the middle of the chalkboard and drawing a circle around it. The class may then search the corpus for synonyms of that particular word and draw lines extending from the word in the middle out to each of the synonyms. Those synonyms then receive their own circle and by doing a synonym string search synonyms of those words can be done as well. Figure 20 shows an example of a word web with synonyms. This can help students visually see the relationship between different words and concepts.


Figure 20. Word web of synonyms for comply.

## Chapter 6: Collocational Knowledge

## Introduction

Collocates are other words' best friends; they hang out together and are often found in each other's company. In more technical terms, corpus linguist Susan Hunston (2002) defines collocation as "the statistical tendency of words to co-occur" (p. 12). Accordingly, the principle of words occurring together is known as collocation and the words that occur together are collocates. The words television, TV, network, and guy are the most frequent words that occur with cable. Thus, one might say cable and television are collocates, or television, TV, network and guy are collocates of cable.

May Fan (2009) studied the use of English collocations by ESL students and found that these students used fewer collocations than native speakers and were unable to collocate words correctly. She also found interference from their native language collocations. An example of this is when Spanish speakers say, "take decisions" based on the translation of "tomar decisions." As a result of her study, she concluded, "Since collocational use is not rule-governed and, in most cases, arbitrary and idiosyncratic, it is important for teachers to raise the awareness of L2 learners to this problematic aspect of language" (Fan, 2009, p. 115). Therefore, because collocations are troublesome and students do not naturally know how to use and learn them, it is essential for them to be taught as part of vocabulary instruction. Corpora help identify collocates, so DDL can be an effective way to increase students' collocational awareness.

Collocations give specific information about how words are used. Noun collocates of verbs indicate the subjects and objects that are used with those verbs. Verb collocates of nouns indicate the properties of those nouns and what they can do. Grammatical collocations, known as colligation (though included together under one category here), help indicate which prepositions
and articles to use with certain words. Semantic and grammatical information can be inferred through specific collocates as well. Upon learning the verb collocates of survey (n) (conduct, find, show, complete, and respond etc.), students might notice that survey is almost only used as the object in a sentence. Frequent noun collocates of sign (v) are contract, agreement, president, bill, treaty, letter, and name. One might infer that the verb sign involves a human subject because only humans can be president and sign names, and is a transitive verb (requires an object).

Though students may not process collocational information in such linguistic terms, studying the patterns of how words interact with each other can help them develop intuitions and inferences to use the words correctly.

## Idioms and other phrases

Perhaps one of the most useful applications of collocational knowledge is with phrasal verbs and other idiomatic expressions. In fact, one significant contribution of corpus linguistics to vocabulary learning is the idea that vocabulary is often more than single words; vocabulary is also idiomatic expressions, set phrases, variable phrases, and phrasal verbs (Folse, 2004). Phrasal verbs are often the bane of non-native English speakers' language learning. They often have nonliteral meanings, can differ by only a single preposition (such as catch up and catch on), and require rote memorization.

Sinclair (1991), was one of the first to introduce the idea that vocabulary is more than a single word. He called it the idiom principle, in contrast to the open choice principle. The idiom principle states that meaning must often be considered in multiword chunks rather than in parts. English language learners who hear the phrase "He needed to man up" might not ever understand what the phrase means if man and $u p$ are considered separately. Learners who have learned to process man up as a phrase follow the idiom principle and are able to comprehend meaning. If
language users are unable to find meaning from larger units during communication, they often fall back on the open-choice principle, which follows the basic grammatical constraints of the language.

Language ambiguity occurs when interpretations from the idiom principle and openchoice principle are both possible (Hunston, 2002). The phrase take someone for a ride could either mean to bring someone along on a leisure drive (open choice) or to be deceived by someone (idiom principle). However the two principles cannot occur simultaneously, and therefore language users must choose which one is more appropriate based on collocation and phraseology. (Hunston, 2002). This might help explain why language learners struggle so much with the phrases of English; they are deciding to consider a phrase as a whole or in parts.

Corpus linguistics research has not only helped us recognize this problem, but also how to solve it. For example, Gardner \& Davies (2007) determined the most frequent phrasal verbs in the British National Corpus so that language learners can focus on structures that they might not recognize naturally as the idiom principle. In order for students to be able to choose meaning correctly from the idiom principle, it is helpful for students to be familiar with some idioms and their format. For example, the fact that grab and bite collocate together is because grab a bite to eat is a common idiomatic expression in English.

Though other vocabulary strategies are useful for remembering vocabulary, inductive DDL methods can help students familiarize themselves with patterns of word use to identify phrases. One difficulty in identifying phrasal verbs is determining if the words on the page should be taken literally or have a figurative meaning. As discussed previously, Sinclair's (1991) idiom principle and open-choice principle can help guide language learners with the latter issue. The idiom principle says to first take meaning as a whole. Therefore, a reader should consider
what hang out means as a whole; if no meaning can be constructed, then the open-choice principle says to look at meaning word by word. Hence, DDL provides additional information that can be helpful in learning idioms and other phrases.

## Mutual information

One final note related to collocations is the idea of relevance (mutual information) versus frequency in conducting collocate searches. Though it is not essential to performing searches, sorting by mutual information can help teachers better navigate the corpus and refine their searches. The default setting in COCA is to sort by "frequency," which gives you the words that occur in the highest actual count. This includes sometimes undesirable results such as $a$, an, the and frequently occurring grammatical words. Hence, another option available is to sort by "relevance," or mutual information, under the sorting and limits section, and to choose "relevance" under the sorting option. By calculating a mutual information score (MI), relevance shows how intimately the words are connected. MI is how frequent each word occurs in the corpus compared to how often they occur near each other through a mathematical formula (for more details see http://corpus.byu.edu/MutualInformation.asp).

Searching by relevance is especially useful because it sorts out the unwanted grammatical terms and can often demonstrate a close relationship between words. On the COCA website it explains, "If you are doing word comparisons, then relevance shows which collocates occur with Word 1 but not Word 2 and vice versa. If you are comparing two sections of the corpus, then relevance shows what words are in Section 1 but not Section 2 and vice versa" (Davies, 2008-). The details of mutual information are linguist-oriented and more complicated to understand, consequently teachers might find it more useful for preparing materials rather than explaining mutual information for student learning.

## Using the corpus

Basic wildcard collocate searches in COCA are fairly straightforward. To search for the most frequent collocates, or neighboring words, simply type the word into the word box and insert a * into the "collocate" query box on the main page by clicking on the word "collocates." Also, if nothing is entered into the collocates box after clicking on the word, COCA will automatically enter a * (Figure 21). The default settings will yield collocates of any part of speech (because of the wildcard *) 4 words before and 4 words after the searched word or phrase. When mutual information is selected for these searches, it should filter out common words like the, an, or $a$.

| display | SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| $\bigcirc$ list Ochart Okwic Ocompare |  | - | CONTEXT | тот |
| SEARCH STRING | 1 | $\square$ | DOOR | 18236 |
| WORD(S) [open].[V*] | 2 | $\square$ | EYES | 6635 |
| COLLOCATES * 4 \% 4 \% | 3 | $\square$ | DOORS | 4047 |
| POS LIST | 4 | $\square$ | MOUTH | 3515 |
| RANDOM SEARCH RESET - | 5 | $\square$ | window | 1573 |
| SECTIONS $\square$ show | 6 | $\square$ | FIRE | 1224 |
| $1)_{\text {IGNORE }}^{\text {I--- }}$ (0) 2 IGNORE |  |  |  |  |

Figure 21. Simple wildcard collocate search for open.

Whereas the default wildcard * will find collocates of any part of speech, collocates of a specific part of speech can be specified by choosing a POS tag. This can be done by placing the cursor into the "collocates" query box and inserting the correct POS tag. If it is unknown, it can be selected from the "POS list" box. For example, to choose only noun collocates for radical, place the cursor in the "collocates" box and select the all nouns tag from the list and [nn*] will appear in the box (Figure 22).


Figure 22. Part of speech collocate search.

To adjust the window of words around the main word, simply choose a number between 0 and 9 in each box after the "collocates" query box. The first number represents the window of words before the main word, and the second number refers to the window of words after the main word (Figure 23).


Figure 23. Search limits for collocate searches.

The "compare" function (radio button) finds collocates for single words present in the first category that are either not present or not as likely in the second category. Consequently, one can compare two words or phrases and their differences in meaning by comparing their collocates. These searches may also be restricted to a single register, which will be discussed in the next chapter.

Collocate searches are one of the most useful and versatile features of COCA. There are many different options that can be pursued to find phrases and collocations. Consider Table 6
showing variations of collocate searches that can be performed in COCA. In the table, "node" refers to the word entered into the "word" search string box, "collocates" refers to what is entered into the "collocate" search string box, and "span" refers to the number of slots around the word on the right or left to look for the collocates (as in Figure 23). "Sort by" and "group by," though not covered here, allow for the user to sort by frequency or relevance and offer a method for grouping the results. Users can modify both under the "sorting and limits" and "options" sections on the left side bar of the screen. An additional feature mentioned in Table 6 is user-created word lists. For more information on this and other features, click on the small question mark to the right of "collocates" or select the appropriate topic from the help bar on the COCA website.

## Table 6

Collocational searches in The Corpus of Contemporary American English. Reproduced from (Davies, 2010)

| NODE | COLLOCATES | $\begin{aligned} & \text { SPAN } \\ & \text { ( L/R ) } \end{aligned}$ | EXPLANATION | SORT BY GROUP BY | EXAMPLES |
| :---: | :---: | :---: | :---: | :---: | :---: |
| laugh.[n*] | * | 5/5 | Any words within five words of the noun laugh | Percentage Collocates | hearty, scornful |
| [thick] | [nn*] | 0/4 | A form of thick followed by a noun | Frequency Collocates | glasses, smoke |
| [look] into | [nn*] | 0/6 | Nouns after a form of look + into | Frequency Collocates | eyes, future |
| [eye] | clos* | 5/5 | Words starting with clos* within five words of a form of eye | Frequency Both words | closed // eye closing // eyes |
| [feel] like | [*Vvg*] | 0/4 | A form of feel followed by a gerund | Frequency Collocates | crying, taking |
| find | time | 0/4 | Find followed by time | Frequency Collocates | time |
| work/job | hard/tough/difficult | 4/0 | Work or job preceded by hard or tough or difficult | Frequency <br> Both words | hard // work tough // job |


| NODE | COLLOCATES | SPAN <br> $(\mathrm{L} / \mathrm{R})$ | EXPLANATION | SORT BY <br> GROUP BY | EXAMPLES |
| :--- | :--- | :---: | :--- | :--- | :--- |
| [=publish] | [n*] | $0 / 4$ | Nouns after a <br> synonym of publish | Frequency <br> Both words | publish // book <br> issue // statement <br> print // money |
| [=expensive] | [[jones:clothes]] | $0 / 5$ | Synonym of expensive <br> followed by a form of <br> a word in the clothes <br> list created by jones | Frequency <br> Both words | expensive / shoes, <br> pricey // shirt |
| [=boy] | [=happy] | $5 / 5$ | Synonym of happy <br> near a synonym of boy | Frequency <br> Both words | happy // child, <br> delighted // boy |

## Applications and sample lessons

The versatility of collocational searches allows for many possibilities and options for teachers to engage their students in DDL activities. Collocates in DDL can be used in a variety of ways in the language classroom. A few examples include teaching new words, teaching definitions, distinguishing between synonyms, and finding phrasal verbs. Days 1 and 2 of the lesson plan "Learning Academic Vocabulary" (Appendix) give brief examples of introducing students to collocations and how to search for them in the corpus. Consider the following case study that demonstrates a simple way to include collocation in vocabulary lessons:

Ms. Kimball's high-intermediate listening/speaking class regularly listens to news clips during class. Before each lesson, she chooses vocabulary words used in the clip and hands a list out to her students. As students listen to the clip, she has them write how the words are used on an activity worksheet. They then indicate what part of speech the word is. To define the words, Ms. Kimball performs collocate searches on COCA in front of the class and has teams of students guess the correct definition as used in the passage, based on collocates. She acts as a guide for the students and corrects their answers. Ms. Kimball uses a simple collocate search, entering the vocabulary word in the query box and an asterisk for the collocate box. Because she
is leading the searches, she can adjust them if needed, such as limiting the window around the node word or sorting by relevance.

As you listen to today's news clips, listen for the following vocabulary words and write as much of the phrase or sentence they are used in that you can. (News Clip) Then, after listening to the passage, try to identify the part of speech for each of the words.

1. threat:
2. baseless:
3. favoring:
4. productivity:
5. tolerate:
6. bullying:
7. behave:
8. target:
9. labor cases:
10. sick leave:

Now write down the collocates for each word (Table 7), and define them in your own words, based upon the class discussion.

Table 7

Word Meaning and Collocates from News Listening Passage

| Word | Collocates | Meaning (your own words) |
| :---: | :--- | :--- |
| 1. threat: | serious, pose, security, <br> nuclear, terrorist, violence, <br> potential | something that maybe will <br> hurt other people |
| 2. baseless: |  |  |
| 3. favoring: |  |  |
| 4. productivity: |  |  |
| 5. bullying: |  |  |
| 6. tolerate: |  |  |
| 7. behave: |  |  |
| 8. target: |  |  |
| 9. labor cases: |  |  |
| 10. sick leave: |  |  |

Now write down any new words you learned that you want to study later:

Ms. Kimball's class activity is simple and focuses on meaning through collocate searches, while introducing students to principles of collocation. Now consider this example for a more advanced group of students, already familiar with the principles of collocation and phrases:

Mr. Duff's advanced reading class has a list of words they review every week. In order to encourage his students to do something more with their weekly word list, each student is assigned one specific word. With their assigned word, students are responsible to find some sort
of phrase or expression where it is commonly used, by searching in COCA. He has students give one-minute presentations on their word, sharing its meaning as learned in class, its meaning in the phrase from COCA (from a dictionary or other sources), and how they searched for it.

During the first few weeks he did the activity, students required help in generating ideas for their searches, and the presentations took a little longer than a minute each. Mr. Duff showed students how to do a basic collocate search and then how to narrow the search window, and look at examples and common patterns from there. Now students in his class are more comfortable in searching the corpus and enjoy learning the different ways the words are used. Table 8 is an example of an overhead Mr. Duff filled out one Friday during the presentations.

Table 8

Vocabulary Words in Collocational Phrases and COCA searches

| Word | Definition (from <br> class) | Phrase | Meaning | Search in <br> COCA |
| :---: | :--- | :--- | :--- | :--- |
| 1. amend (v) | make changes to <br> improve | amend the <br> constitution | make changes to <br> the constitution | amend.[v*] <br> $* *$ |
| 2. challenge <br> (v) | argue a decision or <br> idea | challenge <br> the status <br> quo | argue the way <br> things are done <br> now | challenge.[v <br> $*] * * *$ |
| 3. coherent <br> (adj) | clear and logical <br> communication | coherent <br> thought | thoughts that <br> make sense | coherent.[j*] <br> collocate <br> $\left[n^{*}\right]$ |
| 4. ebb (n) | gradual movement <br> away or lessening | ebb and <br> flow | quick to act <br> violently | ebb.[n*], <br> collocate * <br> $[4,4]$ |

Teachers can also use much of the collocate information to prepare activities and lessons for their students, such as by choosing words for word lists. Word lists that include not only the target vocabulary word but also collocates that it occurs with are more helpful because they
provide the word plus the context. DDL activities can even encourage students to find collocates that occur with vocabulary words outside of class and discover specific uses for the words they are learning. Though many dictionaries are beginning to include collocational and phrasal knowledge, a combination of collocates and example sentences can help students better see the relationship between the words.

Frequency dictionaries can also be a useful way of introducing collocates without having to find them first in COCA. Unlike regular dictionaries, which provide definitions in alphabetical order, frequency dictionaries are in order of the most-used words in a language. An index at the back in alphabetical order makes finding desired words possible. A Frequency Dictionary of Contemporary American English: Word Sketches, Collocates and Thematic Lists (Davies \& Gardner, 2010) is based on the data from COCA and is designed with English-language learners in mind. It is particularly helpful because in addition to frequency information it includes a word's collocations. Many students have difficulties with regular dictionaries, let alone frequency dictionaries, but they are actually quite easy to use when properly instructed. Day 3 of the "Learning Academic Vocabulary" lesson plan (Chapter Appendix) gives a brief example of how to introduce students to frequency dictionaries and how they can be used in the classroom. Incorporating a frequency dictionary is another way to use DDL without having to directly access a corpus. Because of their portable nature, they can be a great source of DDL in technology-restricted classrooms.

## Chapter Appendix:

## Lesson Plan: Learning Academic Vocabulary With Data-driven Learning

Objectives:

- Using COCA and frequency dictionaries students will learn 30 vocabulary words
- Students will list the meaning, part of speech, and collocates associated with the word
- Students will demonstrate an ability to use frequency dictionaries
- Students will demonstrate an ability to use COCA

Words taught: Indicate, individual, interpretation, involved, issues, labor, legal, legislation, major, method, occur, percent, period, policy, principle, procedure, process, required, research, response, role, section, sector, significant, similar, source, specific, structure, theory, variable.

## Instructions for Teachers:

This lesson is divided into three smaller lessons to allow for students to become adjusted to the ideas and tools used here. Each lesson is meant to take 15-20 minutes each day.

The first day is to introduce students to the idea of collocates, synonyms and example sentences. Teachers should also demonstrate the steps in Day 2 activities. The second day students will learn how to perform the searches on their own and find the information in COCA. The third day students will learn how to use the frequency dictionary.

Using a corpus such as COCA for teaching academic vocabulary can be very rewarding. However, students are unfamiliar with what corpora are and many linguistic functions, so you must pave the way and lead them through the process. Many of your students probably use bilingual dictionaries to understand the words. However, even if they do this they will not understand the full meaning of the word without register and collocate information. If students have a hard time understanding the concepts of register and collocation adjust accordingly.

For Day 1 teachers should first walk students through the steps for indicate before working in small groups for the other two words. Day 2 is similar to Day 1. The first page of Day 3 is not a student worksheet but for teachers to use be able to explain collocation to students followed by a worksheet for students to use.

## Day 1: Demonstrate and Introduce (Student worksheet)

## 1. Indicate

- Guess the meaning of the word:

Look at the following examples:
a) Our research does indicate that the public believes that the mission will ultimately succeed.
b) He used it to indicate that we're going to be tough.
c) That would seem to indicate that there wasn't a big risk.
d) But, overall, there's nothing to indicate that we are being harmed.
e) I want you to indicate with your applause who you believe.

Try and guess the meaning of the word by substituting a synonym for indicate.
Guess:

Understanding the word and knowing how to use it will improve with knowing words that are used often with it. These words are called collocates.

Look at the list of collocates when you search for all forms of the word "[indicate]":

1. that
2. results
3. research
4. studies
5. data
6. study
7. findings
8. respondents
9. evidence
10. significant

What do these words describe? What does this tell you about how to use the word indicate?

## 2. Individual

- Guess the meaning of the word:

Look at the following examples:
a) And it's not about a corporation. It's about the individual.
b) What does it mean for an individual?
c) Why wasn't this individual's visa taken away?
d) Well he's clearly a smart individual.
e) Susan was a very disciplined individual.

Try and guess the meaning of the word by substituting a synonym for individual. Guess:

Look at the following examples:
a) And what about individual personalities?
b) The physician in the hospital is paid for each individual service.
c) We have achieved successes with individual species.

This word can be used as a noun and as an adjective. What is the difference in meaning between the two?

Understanding the word and knowing how to use it will improve with knowing words that occur frequently with it. These words are called collocates.
Look at the list of collocates when you search for all forms of the word "[individual]":

1. each
2. rights
3. groups
4. responsibility
5. basis
6. business
7. specific
8. mandate
9. decisions
10. investors

What do these words describe? What does this tell you about how to use the word individual?

## 3. Interpretation

- Guess the meaning of the word:


## Look at the following examples:

a) Their interpretation of situations is so often different from the public.
b) I don't understand what other interpretation there could be.
c) The Pakistanis had a different interpretation of what was going on than we did.
d) My duties involved interpretation of satellite imagery.
e) In "The Interpretation of Dreams", Freud writes that there are two types of dreams.

Try and guess the meaning of the word by substituting a synonym for interpretation.
Guess:

Understanding the word and knowing how to use it will improve with knowing words that occur frequently with it. These words are called collocates.

Look at the list of collocates when you search for all forms of the word "[interpretation]."

1. data
2. results
3. research
4. studies
5. analysis
6. study
7. findings
8. bible
9. evidence
10. constitution

What do these words describe? What does this tell you about how to use the word interpretation?

## Day 2 (Student worksheet)

Students should work in small groups of 2-3 people.

1. Involved

Guess the meaning of the word:

## Look at the following examples:

a) His research interests mostly involved the regulation of genes.
b) Paul had no idea what was involved with the game.
c) There's usually a lot of rage involved.
d) That's not what I've heard. But I don't really get involved anymore.
e) Alcohol was involved in more than 90 percent of hit and run accidents.

Try and guess the meaning of the word:
Follow the steps below to find more example sentences for involved.

1. Go to www.americancorpus.org and click Enter. Make sure you are using an oncampus computer or you will have to register. If you register it's free and you can make unlimited searches.
2. In the word box on the left type the word you are searching for in brackets. E.g.,[involved]
3. Click search.
4. The results will appear on the right. Click on the word in blue to see example sentences.


Find examples of the following words. Make sure that you understand the example sentences that you choose.

1. Find 3 examples of each word.

Write down the best example you find. Write what register it is from: Academic, fiction, magazine, news or spoken. 2. Now find a classmate who has a sample sentence from a different register for the same words. Compare the sentenc

| Example |  |
| :--- | :--- |
| issues | Register |
| labor/labour |  |
| legal |  |
| legislation |  |
| major |  |
| method |  |
| occur |  |
| percent |  |
| period |  |
| policy |  |

## Day 3 (Teacher instructions)

Students will use a frequency dictionary to explore collocate information for words as well as understand what a collocate means. Students will be divided into groups of 3.

Introduce the idea of collocates: Collocates are words that other words "hang out with" or words that are commonly found together. For example in English you can say: have a meeting as in "I have a meeting at 3 pm " but you cannot say "do a meeting" as in "I do a meeting at 3 pm ." Those words don't go together.

Students are probably familiar with the idea of collocates without knowing it. Ask students to fill in the blank of the following phrases:

1. Merry $\qquad$ (Christmas)
2. Happy $\qquad$ (Birthday, New Year)
3. Ride your $\qquad$ (bike, motorcycle)
4. Shake his $\qquad$ (hand)

There is no rule why any of these words belong together. That's a part of leaning words.
A frequency dictionary gives us information about other words a word hangs out with. The dictionary will tell us what the part of speech is, and if the word comes before or after. Look at the following example:

```
147 family n
adj whole, extended, royal, entire, poor, nuclear,
immediate, wealthy, middle-class, low-income noun
member, friend, child, life, .history, support, parent,
.planning verb live, raise, own, visit, gather, feed.,
belong, extend, reunite, flee
147 family \(n\)
adj whole, extended, royal, entire, poor, nuclear, immediate, wealthy, middle-class, low-income noun member, friend, child, life, .history, support, parent, belong, extend, reunite, flee
```

(Source: A Frequency Dictionary of Contemporary American English, Davies \& Gardner, 2010)

This entry tells us that the word family is a noun. Adjectives that occur with the word family are: (adjectives are usually before the word) whole family, extended family, royal family, entire family etc. Nouns that occur with the word (nouns are usually after the word) are family member, family friend, family child, and family life.

Verbs that occur with the word family are live in a family, raise a family, feed the family. (The small dot tells us if the word family comes before or after the collocate if it's not the normal pattern).

To look up words in a frequency dictionary, do not look in alphabetical order like a regular dictionary. You must look alphabetically in the index, and it will give you a number to look up. The numbers show the order of how often the words are used in English.

Therefore collocates are very useful because they tell us how to use a word in relation to other words. This is important information to know because words are used differently in English than in other languages. Assign students to complete the attached worksheet in groups of 3.

## Collocate Activity

1. Divide up the words in your group.
2. With a dictionary or thesaurus write a short definition in English of the word.
3. With the frequency dictionary choose 2-3 important collocates for that word.
4. Write a phrase with the word and collocate together.
5. Share the information with your group.

| Word | Definition in English | Collocates | Phrase |
| :--- | :--- | :--- | :--- |
| Principle | a basic truth or rule | basic, moral, practice, follow, violate | moral principle, violate principle |
| Procedure |  |  |  |
| Process |  |  |  |
| Required |  |  |  |
| Research |  |  |  |
| Response |  |  |  |
| Role |  |  |  |
| Section |  |  |  |
| Sector |  |  |  |
| Significant |  |  |  |
| Similar |  |  |  |
| Source |  |  |  |
| Specific |  |  |  |
| Structure |  |  |  |
| Theory |  |  |  |
| Variable |  |  |  |

## Chapter 7: Register Knowledge

## Introduction

To know a word, one must not only know what it means but also when it is used.
Knowing the meaning of a word is aided by knowing the meaning in addition to the context it belongs in. The idea that words have specific meanings in specific contexts is called register. Nation (2001) describes register knowledge as, "Where, when and how often would we expect to meet this word? Where, when and how often can we expect to use this word?" (p. 27). Some issues of metalinguistic awareness previously discussed have focused on word forms (parts of speech, morphology), and others have focused on word meaning (synonymy, collocations); register focuses on word use. It can be helpful in deciding which synonyms to use, which words to teach, separating multiple meanings of words, and building awareness of pragmatic issues such as language differences between speech and writing.

Like collocational knowledge, register knowledge can help distinguish between different meanings of words. Polysemous words usually have specific meanings of words found in specific contexts, and therefore can be separated by register. Shot used in an academic register could refer to a medical injection, whereas in a fiction or spoken context, it is more likely to refer to a gunshot. The two meanings of the word will differ in collocations, but knowing about register differences can help students know which collocates belong to which context.

Learning about the register of words also allows students to understand pragmatic issues of when to use words. For example, the word gonna appears quite commonly in spoken English, and students may frequently hear it used and know that it means going to. However, in order to use it appropriately, students also need to know that gonna is primarily used in spoken English and hardly ever appears in written, published English, with the exception of quoted dialogue.

This is particularly a relevant issue for those Ferris (2009) calls "ear learners," those who have learned English primarily from speaking it. They often struggle with academic writing because they fail to recognize register differences, and consequently write how they speak. On the other hand, students who are prolific readers might not know that words encountered in text are not always spoken. Without an understanding of register they might receive strange reactions using misappropriate in conversation. Corpus data shows that misappropriate is only found once in spoken English. However, misappropriate is quite common in academic English.

COCA categorizes words into five registers: spoken, fiction, magazine, newspaper and academic. On the interface, they are labeled as sections because the corpus is also divided by time, and years are included underneath the registers. Sub-registers (also called sub-sections in the corpus) can often be more helpful in identifying specific word uses because registers are large and not topic specific. Each register in COCA is divided into several sub-registers based on topic. The magazine register, for example, has several different categories of magazines such as news/opinion, financial, science/technology, social/arts, religion, sports, entertainment, women's/men's, African American, home/health and children. Using the sub-registers to search for frequent words, and collocations can be particularly useful for separating meanings of words. Drive as a verb may mean many things, but in sports magazines ("he hit a thirty yard drive"), it means something different than in financial magazines ("drive up the price"). Though both are within magazine register, the sub-register is where the distinction is made. Therefore, comparing collocates of a word in different registers and sub-registers can demonstrate separate meanings of words. These differences can also be useful in identifying which words to teach for specific content areas.

In the world of English for academic purposes (EAP), registers and sub-registers become especially important in identifying how words are used and how they collocate differently in specific contexts than in other general uses. While a word may still have multiple uses within a single register, sub-registers allow users to narrow down the possible meanings and see more examples of less frequent meanings. As was mentioned in the discussion on frequency, when deciding which words to teach, there are general word lists, academic word lists and word lists for specific contexts. COCA is a great resource for general words to teach and can be a great starting to point to find words for more specific contexts. However, COCA was created to describe general English, and is limited by size for specific topics. Therefore, it should not be the ultimate and only source of EAP or ESP word lists.

Teachers must be aware of the register differences for in-class vocabulary words and help students to understand those differences as well to learn appropriate word use. Such detailed information cannot be gleaned from just one exposure to a word, and can be difficult for students to pick up on their own. Consequently, register should be taught as vocabulary words are recycled through the curriculum in multiple exposures (Schmitt, 2000). Additionally, teachers can build students' register awareness as they examine different kinds of text, fiction, nonfiction, news, etc. in the classroom.

## Using the corpus

There are several different ways to see how a word or its collocates vary in frequency across register. The most basic way is to choose the "chart" radio button at the top of the screen and enter the word or phrase in the "word" query box. This will show the distribution of the word across registers as well as time. To get more specific sub-register information simply click on the
register title above the bar graphs. To see KWIC or example sentences for each sub-register click on the bar (Figure 24).


Figure 24. Chart search comparing neglect across registers.

Another way to see register variation is to choose the "list" radio button and enter the search string in the "word" query box, then check the "show sections" box underneath the search button. This yields similar information but without the bar graph visual (Figure 25).


Figure 25. List search and showing register frequency.

It is also possible to perform basic frequency and collocate searches and specify what register or specific sub-register to search in. The default is set at "ignore" but settings can be specified to search for a word with a KWIC or collocate display from a particular sub-register. The user can choose more than one sub-register as well by holding control on the keyboard while selecting. As an example of choosing certain registers, students writing an academic paper might
want to see example sentences of "however" only in academic writing and therefore choose only the academic register (Figure 26).

```
2010 ACAD AfricanArts A B C Among the many secondary effects of Dakar's new traffic flow is the reality -- however fleeting it may prove -- that much ambulatory ,
2010 ACAD AfricanArts A B C flipping through the pages of Western museum catalogs of African sculpture. n11 here are, however, clear traces of the bieri tradition i
2010 ACAD AfricanArts A B C transferring the reliquary, for procuring skulls, and for performing autopsies. Ironically, however, autopsies were critical to carrying out
2010 ACAD AfricanArts A B C bill is not allowed to leave the continent. The actual bieri head depicted, however, left Africa in the late nineteenth century and has cirs
2010 ACAD AfricanArts A B C FANG (Falgayrettes-Leveau 1991). They have not seen the show or book, however, since there is no copy at the National Library. Ther
2010 ACAD AfricanArts A B C Africa and follows its tributaries to the Diaspora for the theme of the exhibition. However, although he narrows his thesis subject, as ir
2010 ACAD AfricanArts A B C the temporary, the fine from the coarse, the mature from the young. However, it is in the very space of the ephemeral -- the liminal, t
2010 ACAD AmerScholar A B B C for whom Howells was a powerful spokesperson yearned to write of what they saw, however small or everyday, as they saw it. Nothing
2010 ACAD AmerScholar A B C C single day, and partly in the way it makes that idiosyncratic clamor universal. However different the textures of our own lives may be,
2010 ACAD AmerScholar A B C 's no question that it exacerbates the feeling of homelessness. It's hardest, however, for writers. They can't help but be painfully awar
2010 ACAD AmerScholar A B B C . Some writers can't cope with the separation and grow silent. Others, however -- maybe the ones with greater linguistic aptitude -- tr
2010 ACAD AmerScholar A B B C then cook Polish food, and visit Poland every year. I don't, however, wallow in nostalgia and pine for Poland, weeping into my borscht &
2010 ACAD AmerScholar A B C at home on a macro rather than on a micro scale. The local, however, if it is ever to be transformed into a place we call home,
2010 ACAD AmerScholar A B C of health care, or bash the political ambitions of the church. There, however, people assume I'm out of the loop. I live abroad, how
```

Figure 26. KWIC of however in the academic register.

Also, if students are learning the word "dunk.[ $\mathrm{v}^{*}$ ]" in the basketball sense, they might want to search in sports magazines to see example sentences (Figure 27). Searches that focus on specific time periods may be performed in the same manner by selecting years instead of register or sub-register.


Figure 27. Dunk in the sports magazine sub-register.

Finally, it is also possible to compare the usage of a word or phrase between two subregisters or time periods of the corpus. By selecting one sub-register in section one and a different sub-register in section two, the results will show the difference in counts (tokens) between the two for the specified search string. This is possible by first entering the search term into the "word" query bar, then selecting two categories under the "sections" heading (Figure 28). This search can be performed to compare sub-registers of the corpus as well. In order to differentiate between different meanings of a word in sub-registers, enter the search term into the "word" query bar and a term in the "collocate" query bar with appropriate limits (slots on each side). Select two different sub-registers to compare and the minimum frequency or relevance (Figure 28). One caveat is that because of the small size of some sub-registers, comparison collocate searches may not always yield great results. This will depend on the sub-registers being used, the minimum frequency and search limits on either side of the word.


Figure 28. Selecting registers to compare under the sections heading.



Figure 29. Comparing the collocates of two sub-registers.

In addition to being able to compare collocates between registers, the "compare sections"
feature on the "list" screen allows users to compare wildcard search results for phrases. Such a search will show the possible words to fill the blank and how they differ in each register. For example Figure 30 shows the differences between phrases with rest and a preposition, "rest [i*]
[ $\left.n n^{*}\right]$," for magazine and academic registers of the corpus.


Figure 30. Wildcard phrase search comparing two registers.

## Applications and sample lessons

For some students the idea of register may seem very new and unusual. They may have never considered how language is used in different contexts, even in their native language. The "Learning about Register" lesson plan gives examples of how students might be introduced to the idea of situation-specific language, and practice using register knowledge. Once students have a basic understanding of the idea of register, teachers might consider activities like the following.

Ms. Burgin's low-advanced academic listening and speaking class reviews ten words from the Academic Word List (Coxhead, 2000) every week. Students have already learned basic definitions of the words in their reading class, so she teaches her students that the definitions of words might change depending on the situation they are used in. She also explains that some words might only appear in specific sub-registers of academic contexts. While she reviews the words with the students, she has them guess which sub-register of "Academic" each word is more likely to appear in. She gives her students the following worksheet and has them work in small groups.

Each of the following words is from the Academic Word List. For each word, first write the definition you learned in class. Then guess the sub-register of Academic English where it will be the most common. Write the letter of the sub-register next to the word.

Sub-registers are: education, history, geography/social science, law/political science, humanities, philosophy/religion, science and technology, and miscellaneous.

You may use your dictionary if you think it will help. You only have 15 minutes to guess.

1. __fluctuate $\qquad$
2. cease
3. __commence
4. __constitute

| 5. | __area |
| :---: | :---: |
| 6. | ___credit |
| 7. | __appropriate |
| 8. | __deduce |
| 9. | _region |
| 10. | __regulate |

Ms. Burgin gives the groups a limited amount of time in order to force them to think quickly; yet, she gives them sufficient time to discuss any answers they may not agree on as a group. After the time limit has passed, Ms. Burgin leads a class discussion to check their answers as a class. She has students raise their hands about what their group guessed, and she writes their responses on the board. Then, with a computer attached to a projector for COCA (or pre-printed overheads of screen shots), she reveals the answers and shows the frequency of each word in sub-registers. She finds the answers by searching for each word under the "chart" function. When the chart appears she clicks on the column heading for "academic" which reveals the subregister frequency. Ms. Burgin then uses this as a discussion point for each word and why it would be used more in each of these contexts.

If she desires to further the discussion, Ms. Burgin can also show her students how the words' meanings differ in the academic register from other registers, by using either synonyms or collocates. Which she chooses will depend on the word and her students, but with a quick search before class she will ensure the best results. If she choses a synonym search, she can search specific to a register using the "list" function, select academic as one register and another register she would like to compare it to (Figure 31). This will show the synonyms for the word in each
register side by side. To differentiate the definitions by using collocates, she can do a wildcard collocate search on the list function and compare registers (Figure 32).


Figure 31. Synonym search compared in two registers.


Figure 32. Collocate search compared across two registers.

Ms. Burgin uses the activity above to review definitions and develop her students'
metalinguistic awareness. In reviewing the definitions as a class, she allows students to use
dictionaries if needed and during the activity she encourages them to take notes. After students
are familiar with the activity it goes very quickly, takes limited preparation on her part, and helps students learn to develop language intuitions based on their experiences so far.

Mr. Owens likes to use a different activity to review register and collocational knowledge in his high-advanced reading class. His students already understand the basic concept of collocates and register differences. His goal is for his students to be able to get more exposure to their vocabulary words and see how words can be used differently depending on register and collocates. He has the students (in groups or alone) use the "compare sections" function of the corpus to show differences in collocates in different sub-registers. With the aid of COCA and a good dictionary, Mr. Owens helps his students develop the ability to see the small differences in meaning based on context. This practice greatly enhances their reading skills and ability to understand words in context. In order to complete the activity, he has students fill out a worksheet that corresponds to the activity (Chapter Appendix: Register in Sub-registers).

## Chapter Appendix:

## Lesson Plan: Learning about Register (Academic English)

Goal: Help students understand some vocabulary differences between academic English and spoken English.

## Learning Outcomes:

- Students will demonstrate an understanding of how language changes in context
- Students will recognize and communicate some differences between academic and spoken English
- Students will know that COCA can demonstrate differences in register


## Materials:

computer attached to projector OR overhead projector with preprinted slides attached worksheet of Academic or Spoken?

## Procedures:

1. (10 min) Introduction: Write the following exercise on the board or overhead. Have students individually match each phrase to the person they would say it to.

Which phrase would you say to which person?
a. Hey, man, give me a hand, will ya? a. professor
b. Would you please help me with my homework?
b. roommate
c. Sir, I was wondering if you might be able to help me answer a question.
c. tutor

Discuss as a class some ways that students have noticed language used differently in different situations.
2. (15 min) Instruction: Demonstrate to students that academic English is very different from English spoken every day. Give them a demonstration with the word, aversion. Ask if any of them have heard of it before. Search for it in COCA (www.americancorpus.org) using the "chart" function and point out that it's very infrequent in spoken English compared to academic English. Do a second search with the word wow, using the "chart" function, and show how common it is in spoken English but not academic English. Explain the following differences about vocabulary in academic English (though there are grammatical differences as well):

- There is an academic word list of words that commonly occur in academic English and some basic English phrases and expressions that will almost never occur in academic English
- Some words are in the middle. They're not basic English words but not academic words either.
- Many words are used in both places, but are more commonly used in one than the other
- Academic words use a lot of Greek and Latin morphemes
- Some words have a different meaning when used in academic English compared to other registers. (e.g., assemble, he assembled the bookshelf vs. the committee assembled regularly).

3. (20 min) Practice: In small groups or pairs, have students complete the worksheet Academic or Spoken. Make sure students understand that the do not have to understand the whole sentence but they are trying to make their best guess. When students finish go over answers as a class and then have students answer the last two questions at the bottom of the page. Answer key: 1-S, A; 2-S, A; 3-A, S, 4-S, A; 5-S, A; 6-S, A; 7-A, S; 8-S,A; 9-A, S; 10- A, S
4. (10 min) Review: Review as a class some of the differences students noticed from the activity. They should be similar to the explanation given at the beginning of class. Tell students that differences are called register and show them how to search for a word using the chart function to see where it is more common. Show them that clicking on the bar graph gives the example sentences. Make sure that you don't expect them to memorize this information but only to become familiarize themselves with it.
5. Homework: If students have computer access, such as at home, the library or in a computer lab at school, consider giving them homework to look up their weekly vocabulary words and write where they are used more common.

## Academic or Spoken?

All of the following words can be used in academic or spoken English. There are two example sentences for each word and a number indicating how often that word occurs in academic or spoken English. Based on the number and the example sentence make your best guess with an A for academic and an S for spoken English.

1. compulsory

67, First of all, I think we should clarify the meaning of terms compulsory and mandatory that get thrown around.
__658, But it was not until 1896 that it became compulsory for archaeological excavations to consult the Mexican government before starting excavations.
2. compel
_695, We can compel the courts to respond to the needs of the citizens.
2167, I do not wish to imply or compel others to infer that the relationship between imagery and activism is symbiotic
3. responsibility

13899, Responsibility for the correctness of any given piece of writing should fall mainly on the student not the teacher. _ 8845, People who work hard and meet their responsibilities get ahead.

## 4. dependent

1069, And certainly a heart transplant was an option, but you're dependent on someone else's misfortune, and you can't pull a heart transplant off

6863, Indeed, the demonstration of benefits may be highly dependent on which two treatments are being compared.

## 5. sidewalk

584, If you just walk down this sidewalk, there's one right about ten yards from me covered with a sheet.
__346, The participants also jumped on asphalt, carpet, cement sidewalk, grass, and the floor of a gymnasium.

## 6. sunrise

__179, That's the first sunrise that I was able to take a picture of.
__294, In landscape photography the best time to shoot great images is at sunrise and sunset.

## 7. courage

1257 , He argued that men must possess a stout courage to exercise their constitutional rights. __2288, But when you try to talk about courage with them, they sort of deny that it exists

## 8. convenient

_ 466, Think about the timing of that statement; that's somewhat convenient.

1013, However, in remote locations where there is no convenient access to electrical power distribution, stand-alone PV systems become very attractive
9. designed (v)
__ 14293 These results demonstrate that because state-based models are designed to create narrow ranges in the classification variable ___ 3647 The megastar ad is designed for, you know, the era of John Stewart and YouTube.
10. isolated (adj)

2092, Wufong was treated as a reference station, because it represents an isolated rural area in the Taichung metropolis.
__863, But in this case, right now, this is not an isolated incident.
Write some of the things you noticed about the spoken sentences:

Write some of the things you noticed about the academic sentences:

## Activity: Register in Sub-registers

Goal: Students will learn that vocabulary meanings can change depending on the context of the word.
Directions:

1. Choose two sub-registers that you are interested in such as "Academic: Education" and "Academic: History."
2. In COCA compare the collocates of a word in the two sections. Limit the range of your collocates to 2 on each side.
3. Look at example sentences of the collocates and write down a few that you think are related to the same meaning.
4. Write an example phrase of how a collocate and the word are used together.
5. Choose a definition from the dictionary that matches the definition of how the collocates are used.
6. An example is done for you using "Academic: Education" and "Academic: History"

Sub-registers used: $\qquad$
$\qquad$
Unit vocabulary words:
analysis (n)
available (adj)
approach (n)
assume (v)
authority (n) benefit ( n )
concept (n) context (n)

| Word | Collocates of Subregister 1 | Example phrases | Possible definition | Collocates of Subregister 2 | Example Phrases | Possible definition |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| access | student, postsecondary, services appropriate | every high school student has access to a computer 24/7 | ability to use or benefit from something | foreign, land, markets, policy, | the country should allow for easier access to land and property | a way to obtain something |

1. 
```
3.
4 .

\section*{Chapter 8: Extension (Autonomous Learning)}

This chapter provides some additional information about how COCA and DDL can be employed in language teaching. It includes adaptions of individualized DDL instruction, case studies, additional corpus-based resources that are useful for teaching DDL, and guides for troubleshooting the corpus.

\section*{Individual Language Learning}

Though the primary focus of this book is teacher-guided DDL activities, once students have an adequate understanding of vocabulary principles and how to use COCA, it can be used for autonomous and individualized study. However, students must first see the validity and utility of the corpora in class, so they do not lose motivation and become frustrated using it. In addition, teachers must realize their goal is not to train corpus experts, but to help students learn basic principles through an inductive process of language learning. In DDL, the teacher still maintains the role of expert but is gradually pushing and building the students to become self-directed. For students to learn vocabulary individually through DDL, teachers must first guide them through the process and give them adequate examples to follow. Class projects, homework assignments, and individual tutoring can be used to teach students to gradually explore on their own. In pushing towards data-driven learning, the teacher is placing more of the responsibility for vocabulary where it belongs--on the learner.

Because language classrooms are full of mixtures of different personalities, learning styles and language abilities, teachers must first be aware of how individuals will feel about and adapt to corpus-based instruction. For the technology-savvy students, DDL may seem like a dream come true, especially if they are at an institution that provides easy technology access. However, these students should be warned that technology is not perfect, and does not hold all
the answers. Corpora are only useful when human interpretations and intuitions are applied. For the technology-anxious students, DDL may cause some stress. Teachers should consider gradually introducing corpora in classes and heavily emphasizing applications to show their usefulness. Working in pairs, groups, or as a whole-class may reduce some of the anxiety until students begin to feel comfortable. Also, in this situation, or if technology is unavailable, teachers might consider using corpus-based resources more than the actual corpus or preparing their vocabulary lessons based on their own findings in the corpus. Finally, students studying English for specific purposes (ESP) can use DDL to identify word lists and topics necessary for their studies. Students studying English for general purposes can make an effort to study the most frequent words in English generated from corpora or found in frequency dictionaries.

After extensive classroom practice, students must learn to apply corpus methods to their own work, particularly to address their questions and in error-correction of their own papers. This may be challenging because they often struggle to recognize their errors and to know how to fix them. As teachers meet with students individually and give feedback on papers they can help by marking the errors and directing students to the corpus. In this way teachers can help students become involved in DDL with their individual learning and show how vocabularyrelated issues can be identified and resolved in a personal way. Then students can learn to find appropriate solutions on their own and become more independent language learners. The following case studies demonstrate how that is possible.

\section*{Case studies}

Pablo \({ }^{1}\) is a high-advanced English language learner. He is in the final level of intensive English study before attending an American university next semester. The following is an excerpt

\footnotetext{
\({ }^{1}\) Names have been changed
}
of a paper written for his ESL writing classes. He met with an ESL teacher, Mrs. Peterson \({ }^{1}\) after turning in the first draft of his paper. The following is an excerpt from his paper. Specific vocabulary errors have been underlined.

In this current second decade of the twenty-first century, the human race is exhibiting symptoms of how it has been tremendously affected by these changes and discoveries. These sometimes dangerous alterations in the behavior of both society and individuals have two common causes: personal computers and the Internet. The evolution of technology and the economy has allowed increasingly ease of access and affordability to the World Wide Web.

Pablo's vocabulary errors in this paragraph are influenced by a lack of metalinguistic awareness such as specific definitions, collocations, context, and morphological forms. To help him understand what the errors are and how to correct them, Mrs. Peterson asks Pablo to perform a few searches in COCA. Her role as the teacher is to help him identify what the error is and to find alternative options by using corpora, thus helping him discover his own solution. Upon searching for "[exhibit] symptoms" and looking at the examples they discover that this is a literal meaning of symptoms regarding health and illness. In order to find a more appropriate word they search for "[=exhibit] symptoms" to find synonyms of exhibit that collocate with symptoms. Three other options appear besides exhibit: show, display and demonstrate. Because demonstrate only occurs once with symptoms they ignore it and Pablo chooses between show and display by looking at the examples and what he likes better.

Pablo's first error is about word choice and specific meaning use, but his second vocabulary error is about the wrong word form and the wrong morpheme. This is a situation where metalinguistic knowledge taught to the students can be helpful. Mrs. Peterson has already taught Pablo about inflectional and derivational morphemes she tells him that he has the wrong form of the word. Together they identify the root of the word and search in COCA for "increas*" which will yield all forms. Upon looking at the results "increase, increased, increasingly, increases, etc.," Pablo will try to identify the correct form. If he cannot correctly select "increased" or "increasing" as the correct form, together they can try a new search with more context such as "increas* ease."

Later, after having received feedback in this manner several times with Mrs. Peterson, Pablo began to use the corpus on his own. When asked about it he commented, "I use [the corpus] a lot, actually. I always use it to find collocates and the most accurate way to use a word [or] phrase. Also, sometimes I come up with phrases I heard from movies or TV shows and I check whether they are academic." The use of data-driven learning, for Pablo, has helped him become a more autonomous learner.

Carlos \({ }^{2}\) is also a high-advanced ESL student, preparing to enter graduate school in the United States. In his content-based university preparation program he regularly meets with a tutor to review key concepts. In a tutoring session one week he wrote the following (error marked):

The Democratic and Republican parties are two of the oldest political parties in the world. The Democratic and Republican parties are the result of ideologists that people

\footnotetext{
\({ }^{2}\) Names have been changed
}
such as Thomas Jefferson and Alexander Hamilton had when George Washington established his cabinet.

Upon reviewing his paragraph, Mr. McArthur \({ }^{2}\), his tutor, marked the error and asked him if he understood the problem. Carlos admitted that he knew it wasn't the right form but was confused between the different forms. Mr. McArthur asked Carlos to identify what he thought the base of the word was, which for Carlos was idea. Together they entered "idea*" into COCA with a minimum frequency of 10 (to avoid infrequent usages) and looked at the results. None of the forms looked like the one Carlos used in his paper or the one he wanted so Mr. McArthur encouraged him to choose a new base form. He chose ideol and entered "ideol*" into the corpus. His chosen form ideologists appeared so they looked at the other results as possible solutions starting at the top. Carlos was unsure of what the word ideology, the top result, meant, so he and Mr. McArthur searched the corpus for *ology based on Mr. McArthur's suggestion and compared the resulting technology, psychology, biology etc. Based on the results he guessed that *ology meant the study or pursuit of something. Mr. McArthur then asked him what the word he had chosen, ideologists meant. A resulting search of *ists helped Carols see that both ideology and ideologists are nouns but one is the beliefs and study and the other are the people who study and believe it.

What Mr. McArthur did with Carlos may seem like a long process to correct one error; however, because both were already familiar with COCA, it only took about seven minutes of Carlos' tutoring. Through the process he was able to better understand the meaning of word endings in English and how he can find the correct word on his own, as he is writing papers. Though he still required the assistance from a teacher, Carlos corrected his own error and in the
process because more autonomous. Hopefully, Carlos will remember the process and be able to repeat it in the future the next time he confuses two related forms of a word.

\section*{Troubleshooting COCA searches}

The previously detailed features of COCA and search instructions in chapters 2-7 include many common and useful parts of COCA's interface for language teachers and students. However, there are many other COCA features that can provide helpful information, such as multiple slot searches, customized word lists, combining features together, as well as more complex variations on the basic searches. Understanding the sorting and minimum frequency features may also be helpful, especially when troubleshooting a search that doesn't turn out quite right. However, in order to not overwhelm new users to the corpus, those features will not be addressed here. Though the introduction to collocate knowledge provides a brief overview of relevance and frequency, those who are interested in more detailed information should look in the help and tour sections of COCA.

Unfortunately, even with limiting the information in this booklet to basic information, often times the unexpected happens and results come out different than expected. The following questions and answers are an attempt to help troubleshoot possible search problems when using The Corpus of Contemporary American English.

\section*{Troubleshooting:}
1. What happens when I did the search I thought I needed to and get the wrong results?

There are several possible explanations here.
a. You may need to be more specific or more general in your search. Try to specify the part of speech for words (word.[n*]) that may be ambiguous, or search for a specific
part of speech collocate. On the other hand, try to give less information such as using a generic wildcard.
b. You may need to specify a register or sub-register.
c. You may need to sort by relevance instead of frequency (Figure 33). For more on this, select the question mark to the right of "sort by" and see the section titled "Mutual information" in chapter 6.


Figure 33. Selecting frequency or relevance settings.
d. You may need to set a minimum frequency ( 10 is good) or a minimum mutual information score (2 or 3 is good).
e. You may have spelled something incorrectly or attached the wrong part of speech tags.
f. You may have selected the wrong display option at the top of the screen. Most searches are done in "list."
g. You may need to narrow or broaden your window on the collocate search. The default is set at 4 on each side. Try making this bigger or smaller or only looking on one side of the word.
h. Try resetting the corpus and entering your results again.
i. The corpus may have a hiccup. This is not usually the case, but if it is, report it on the website.
2. What happens when I get no results at all?
a. You may need to \(\log\) in (after 10 queries).
b. Your phrase is too long; you may only have 10 slots.
c. You may have too many wildcards in your search; try and give some more specific information in one of the slots.
d. The corpus does not have any examples of your phrase (it is only is a sample, remember).
e. You may have spelled something incorrectly.
f. You may have set the minimum frequency or mutual information too high.
g. You are not connected to the Internet.

For more information on features of the corpus and how to use it please see (Davies, 2008-), (Davies, 2009) and (Davies, 2010).

\section*{Other Corpus-based Vocabulary Resources}

DDL, as described in this book, has generally included students interacting directly with language patterns and then making inferences. In this way, DDL can enhance students' analytical abilities of the language and their awareness of language around them. Instructions have primarily directed teachers to demonstrate the corpus in class or have students use the corpus in class or for homework. Often times, technology is not sufficiently available. Not all classrooms have the ability to use computers connected to the Internet on a projection screen, nor do all schools have access to a computer lab for their students. In that case, teachers might consider performing COCA searches before class, taking a screen shot of the results and printing them to
an overhead. Printouts of such data, or concordance lines might also be more practical if computer labs are not readily available. Other DDL alternatives to directly using COCA in class might include other corpus-based resources.

Corpus-based resources are not the same as DDL because they do not allow the students to interact as much with the data directly. However, they can still be very useful for vocabulary learning and language study because they are based on real data. The following are several corpus-based resources that teachers might find useful in teaching vocabulary skills and metalinguistic awareness to their students. These are just a sampling of many corpus-based or corpus-informed resources available.

A Frequency Dictionary of Contemporary American English Word Sketches, Collocates and Thematic Lists by Mark Davies and Dee Gardner (2010) is a DDL, corpus-based resource to help students develop a sense for collocations and frequency. It also contains part of speech information for words and collocates and specialized word lists.

The Oxford American Dictionary Vocabulary Builder by Keith Folse (2011) teaches basic dictionary skills, spelling, pronunciation, parts of speech, collocations, and register for vocabulary skills. It is based on the Oxford University Press corpus. Though it is not DDL it is a great resource for introducing and teaching aspects of vocabulary through specific activities.

Using Corpora in the Language Learning Classroom Corpus Linguistics for Teachers by Gena R. Bennett (2010) gives a good overview of corpus linguistics for non-experts. The author clearly gives a brief overview of the issues, corpus resources and sample activities including guides to creating your own corpora.

Finally, the Touchstone (McCarthy, McCarten, \& Sandiford, 2006) series of integrated skills textbooks is based on the Cambridge international corpus. One small advantage it has over
other textbooks is being corpus based, and it includes collocational knowledge and informal exercises for students.

\section*{Chapter 9: Teacher Feedback}

This book is meant to benefit a variety of ESL teachers. It is for those who see a need and desire to improve vocabulary teaching in the classroom, but who require more knowledge to do so. It is also for those teachers who have heard of corpora or DDL and can see their usefulness but are unaware of how to incorporate them in the language classroom. Also, there might be teachers who stumbled across COCA but are curious as to how it can apply to them. Therefore, the information presented here would be incomplete without considerations and feedback from those it is meant to benefit.

Three ESL teachers of various backgrounds were surveyed to judge the usefulness of this book and how well it accomplished their goals. They were asked various questions about their teaching background, previous knowledge of corpora, vocabulary, and DDL, as well as their opinions on how well this book met its objectives. The original survey can be viewed in the Chapter Appendix.

All of the teachers surveyed had ESL training, at least at a graduate certificate level. In their training, all of the teachers had been introduced to the idea of corpora and had basic training in vocabulary instruction. Respondents commented that much of the information presented in this book was new to them, particularly information on what it means to know a word and how to use corpora. Because of teachers' connections to BYU, where COCA was developed, all were familiar with COCA, but many were unaware how to use it in the language classroom. One respondent said, "I knew about COCA and I've used it before, but not from the EFL/ESL student perspective. I knew that effective vocabulary teaching requires many examples of the words in contexts." Another respondent said, "I knew that corpora could be used to teach vocabulary, but I just didn't know how to use it properly."

Regarding what was most useful to them in the book, all participants responded that the lesson plans and examples helped them the most. Respondents also mentioned that the detailed, specific, instructions on how to use COCA were helpful for them in looking to apply COCA to the classroom. Respondents said that the demonstrations better helped them understand the explanations of vocabulary topics and gave examples to follow. In the words of one respondent, "The applications, case studies, and sample lesson plans were treasures."

All survey participants responded favorably in their desire to apply the concepts and lessons introduced in this book to their own language classroom, though perhaps not as much for the Elementary ESL teacher in her context. She said, "I may not be able to apply many of these methods due to the level I teach (elementary), but I will certainly retain the approaches to analyzing and teaching vocabulary." In fact, this teacher's hesitation to apply DDL methods with her elementary students was the only concern expressed about the content and methods in this book. She said, "This is certainly aimed for intermediate to upper level students and using corpora will require a level of maturity or analytical skills. Even though multiple contexts are advantageous for high-beginning students, I'm hesitant to search corpora since the results seem broad and unpredictable. However, DDL may not be necessary or appropriate for beginning levels (or low maturity levels) so it's only a consideration of mine." Though some studies have been done teaching children how to use concordance lines, the teacher brings up valid points concerning DDL in elementary classrooms. Her comments allow room for speculation and future research on how the principles taught in this book can also be applied to children, and if, indeed, that is necessary or possible. However she later said that when she returns to adult education, "I hope to use corpora to provide vocabulary context for my advanced students. I intend to use these exact methods if/when I teach older and more advanced students."

The other two respondents are already ESL teachers in adult contexts and showed interest in applying the principles and lessons of this book. One of the other respondents said, "Because of the clear sample lessons given in this document, I want to try them when teaching vocabulary." The other respondent agreed to use the methods in her classroom, but desired to see that they worked first. She had already heard of DDL methods before in language classrooms and had not previously tried them. She replied, "Before, I would say that it was 'Very Unlikely' that I would use these ideas and concepts, because I didn't understand how I would use this information, and I didn't want to take the time to figure it out. Now I am 'Somewhat Likely' to use it because of this boost in knowledge that I have. I am not quite 'Likely' or 'Very Likely' because I still haven't used this information (I didn't do any sample searches while I was reading through this document). Once I have used it and seen how it works first hand, I would say that I would be 'Very Likely' to use this information in my teaching."

Though some of the respondents had heard of DDL in their TESOL training, none of the respondents had even attempted using COCA in their language teaching. Therefore, many of the ideas and applications presented were new to them, particularly the concepts of DDL. One respondent replied, "[Before reading this book] I didn't know how to do a variety of searches to accomplish the variety of approaches to vocabulary learning...I didn't know data-driven learning could be an effective approach, or how to incorporate DDL in my classroom." Another participant said, " A lot of information on how to use corpora for data-driven learning was new to me. And I didn't know that corpora could be used to teach about synonymy and morphological knowledge."

Finally, respondents were asked how well this book was able to accomplish its purpose of instructing ESL/EFL teachers how to better teach vocabulary using data-driven learning. All
respondents completely agreed that it did. "It accomplished the purpose in full. It not only proved to me the possibilities and advantages of DDL, but it motivated me to enhance my vocabulary teaching by using corpora and DDL approaches," said one respondent. Another said, "I feel like the purpose was met very well. I now have a better idea of what can be taught using data-driven learning, and I actually feel like it is something that is doable. It was met quite well because the instructions were clear and it gave sufficient information about teaching vocabulary."

In summary, the ESL teachers who previewed this book found it helpful and useful to their ability to teach vocabulary and methods available. The evaluation of this resource book by these teachers shows that the book fulfills its purpose and is useful to ESL teachers. Like these teachers, there are many more who have heard of corpora and DDL but are unsure how to apply it in the classroom. Hopefully, the methods and lessons that are presented in this book are just as helpful to them as they were to the three teachers who responded to the survey.

\section*{Author's note:}

The previous comments and results are taken from a survey of teachers asked to evaluate how useful this resource has been in their understanding of vocabulary teaching and data-driven learning (Appendix). If you have used the information, practices, lesson plans or ideas presented in this book please contribute to its continual improvement with your own feedback by taking this survey or emailing the author.

\section*{Chapter Appendix:}

\title{
Evaluation
}

\section*{Survey Questions:}

Background information
Current position:
Current courses teaching:
How many years have you taught ESL/EFL?
What formal ESL training have you received?
BA major, BA minor, certificate, \(\mathrm{MA}, \mathrm{PhD}\), on the job training
How many foreign languages do you speak?
Vocabulary information
Have you every received training about vocabulary?
If yes, what?
Before reading this document were you familiar with the term metalinguistic awareness?
What information about vocabulary in this document was new to you?
What information did you already know?
What information was more or less useful?
Corpora and DDL
Before reading this document were you familiar with the term data-driven learning?
If yes, where?
Had you heard of the Corpus of Contemporary American English?
If yes, where?
Had you ever used it in teaching? (Never, once a month, two or three times...daily)

What information about Corpora and data-driven learning was new to you?
What did you already know?
What information was more or less useful?
Purpose of the document
The purpose of this document was to instruct ESL/EFL teachers how to better teach vocabulary using data-driven learning. How well do you feel like that purpose was met? Do you feel more prepared to teach vocabulary and use DDL after reading this?

How likely are you to do so?
What final comments do you have about this document?
Would you interested in being notified about more information on these topics in the future?

\section*{References}

Aijmer, K. E. (2009). Corpora and Language Teaching. Amsterdam: John Benjamins. Balunda, S. A. (2009). Teaching academic vocabulary with corpora: Student perceptions of data-driven learning. . (Unpublished MA Thesis). Master of Arts, IUPUI, Indianapolis.
Basanta, C. P., \& Martin, M. E. R. (2007). The application of data-driven learning to a small scale corpus: Using film transcripts for teaching conversational skills. Language and computers: Studies in practical lingusitics, 61, 141-158.
Bennett, G. R. (2010). Using corpora in the language learning classroom: Corpus linguistics for teachers. Ann Arbor: University of Michigan Press.
Boulton, A. (2009a). Data-driven learning: Reasonable fears and rational reassurance. Indian Journal of Applied Linguistics, 35(1), 81-106.
Boulton, A. (2009b). Testing the limits of data-driven learning: language proficiency and training. ReCALL, 21, 37. doi: 10.1017/s0958344009000068
Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. Language Learning, 60, 534-572. doi: 10.1111/j.1467-9922.2010.00566.x
Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. ReCALL, 17(01). doi: 10.1017/s0958344005000510
Campoy-Cubillo, M., Belles-Fortuno, B., \& Gea-Valor, M. L. (Eds.). (2010). Corpus-based approaches to English language teaching. London: Continuum.
Chujo, K., Anthony, L., \& Oghigian, K. (2009). DDL for the EFL classroom: Effective uses of a Japanese-English parallel corpus and the development of a learner-friendly, online parallel concordancer. Paper presented at the Corpus Linguistics Conference, University of Liverpool, UK.
Clifton, J., \& Phillips, D. (2006). Ensuring high surrender value for corporate clients and increasing the authority of the language instructor: The dividents of a data-driven lexical approach to ESP. The Journal of Language for International Business, 17(2), 72-81.
Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? System, 25, 301-315.
Coxhead, A. (2000). A new academic word list TESOL Quarterly, 34, 213-238.
Cresswell, A. (2007). Getting to 'know' connectors? Evaluating data-driven learning in a writing skills course. In E. Hidalgo, L. Quereda \& J. Santana (Eds.), Corpora in the foreign language classroom (pp. 267-287). Amsterdam: Rodopi.
Davies, M. (2008-). The Corpus of Contemporary American English (COCA): 410+ million words, 1990-present. . Available online at http://www.americancorpus.org.
Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (19902008+): Design, architecture, and linguistic insights. International Journal of Corpus Linguistics, 14, 159-190. doi: 10.1075/ijcl.14.2.02dav
Davies, M. (2010). Semantically-based, learner-oriented queries with the 400+ million word Corpus of Contemporary American Englis. In S. Gozdz-Roszkowski (Ed.), Lodz Studies in Language. Frankfurt: Peter Lang.
Davies, M., \& Gardner, D. (2010). A Frequency Dictionary of Contemporary American English: Word Sketches, Collocates and Thematic Lists. London: Routledge.
Fan, M. (2009). An exploratory study of collocational use by ESL students - A task based approach. System, 37, 110-123. doi: 10.1016/j.system.2008.06.004

Ferris, D. (2009). Teaching college writing to diverse students populations. Ann Arbor: University of Michigan Press.
Flowerdew, J. (2009). Corpora in language teaching. In M. H. Long \& C. J. Doughty (Eds.), The Handbook of Language Teaching (pp. 327-350). Oxford, UK: Wiley-Blackwell.
Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. International Journal of Corpus Linguistics, 14, 393-417. doi: 10.1075/ijcl.14.3.05flo
Folse, K. (2004). Vocabulary myths: Applying second language research to classroom teaching. Ann Arbor: University of Michigan Press.
Folse, K. (2011). Oxford American Dictionary Vocabulary Builder. Oxford: Oxford University Press.
Gardner, D. (2008). Vocabulary recycling in children's authentic reading materials: A corpusbased investigation of narrow reading. Reading in a Foreign Language, 20, 92-122.
Gardner, D., \& Davies, M. (2007). Pointing out frequent phrasal verbs: a corpus-based analysis. TESOL Quarterly, 41, 339-359.
Ghadessy, M., Henry, A., \& Roseberry, R. L. (Eds.). (2001). Small corpus studies and ELT: theory and practice. Amsterdam: John Benjamins.
Hadley, G. (2002). An introduction To data-driven learning. RELC Journal, 33, 99-124. doi: 10.1177/003368820203300205

Henry, A. (2007). Evaluating language learners' response to web-based, data-driven, genre teaching materials. English for Specific Purposes, 26, 462-484. doi: 10.1016/j.esp.2007.01.003

Hunston, S. (2002). Corpora in applied linguistics. Cambridge: Cambridge University Press.
Johansson, S. (2009). Some thoughts on corpora and second-language acquisition. In K. Aijmer (Ed.), Corpora and Language Teaching. Philadelphia: John Benjamins.
Johns, T. (1991). "Should you be persuaded": two samples of data-driven learning materials. English Language Research Journal, 4, 1-16.
Koda, K. (2000). Cross-linguistic variations in L2 morphological awareness. Applied Psycholinguistics, 21, 297-320.
Liu, D. (2010). Is it a chief, main, major, primary, or principal concern? A corpus-based behavioral profile study of the near-synonyms. International Journal of Corpus Linguistics, 15, 56-87. doi: 10.1075/ijcl.15.1.03liu
Markovic, J. (2002). The influence of Greek and Latin on the lexicon of narrative and expoistory text collections in upper-elemenatry education: Implications for extensive reading. MA, Brigham Young University, Provo, UT.
McCarthy, M. (2008). Acessing and interpreting corpus information in the teacher education context. Language Teaching, 41, 563-574.
McCarthy, M., McCarten, J., \& Sandiford, H. (2006). Touchstone. Cambridge: Cambridge University Press.
McEnery, T., \& Wilson, A. (2001). Corpus Linguistics (2nd ed.): Edinburgh University Press.
Moon, R. (2007). Sinclair, lexicography, and the cobuild project. International Journal of Corpus Linguistics, 12, 159-181.
Nation, I. S. P. (2001). Learning vocabulary in another language. Cambrindge: Cambridge University Press.
O'Keeffe, A., McCarthy, M., \& Carter, R. (2007). From corpus to classroom: language use and langauge teaching. Cambridge: Cambridge University Press.

Pérez-Paredes, P., \& Alcaraz-Calero, J. M. (2009). Developing annotation solutions for online Data Driven Learning. ReCALL, 21, 55. doi: 10.1017/s0958344009000093
Read, J. (2010). Methodological issues in research on vocabulary learning. In Y.-N. Leung (Ed.), Proceedings from the 19th International Symposium on English Teaching, Taiwan (pp. 63-73).
Romer, U. (2008). Corpora and language teaching. In A. Ludeling \& M. Kyto (Eds.), Corpus linguistics: An international handbook (Vol. 1, pp. 112-130). Berlin: Mouton de Gruyter.
Romer, U. (2009). Corpus research and practice: What help do teachers need and what can we offer? In K. Aijmer (Ed.), Corpora and Language Teaching. Philadelphia: John Benjamins.
Schmitt, N. (2000). Vocabulary in language teaching. Cambridge, U.K.: Cambridge University Press.
Sealey, A., \& Thompson, P. (2007). Corpus, concordance, classification: Young learners in the L1 classroom. Language Awareness, 16, 208-216.
Sinclair, J. M. (1991). Corpus, Concordance, Collocation Oxford: Oxford University Press.
Sinclair, J. M. (2004). How to use corpora in language teaching. Amsterdam: John Benjamins.
Sinclair, J. M., \& Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter \& M. McCarthy (Eds.), Vocabulary and Language Teaching. London: Longman.
Wang, J., Liang, S., \& Ge, G. (2008). Establishment of a Medical Academic Word List is . English for Specific Purposes, 27(4), 442-458. doi: 10.1016/j.esp.2008.05.003
Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. English for Specific Purposes, 28, 170-182. doi: 10.1016/j.esp.2009.04.001

West, M. (1953). A general service list of English words. London, New York: Longmans.
Widdowson, H. G. (1998). Context, community and authentic language. TESOL Quarterly, 12, 705-716.```

