Corpora and English Language Teaching: Pedagogy and Practical Applications for Data-Driven Learning

Michael Lessard-Clouston, Biola University, California
Tracey Chang, Bethesda University of California

Introduction

Although corpus linguistics has become quite influential in English language teaching in recent years, the actual use of corpora (large, searchable collections of real language, electronic texts) by practicing teachers still seems to be fairly limited. In their introduction to data-driven learning (DDL) in the classroom, Gilquin and Granger (2010) declare, for example, "one reason for not doing DDL might simply be that the teacher does not know enough about corpora and the possibility of using corpora in the classroom" (p. 366). This article aims to help fill this gap, by assisting teachers in understanding potential uses of corpus linguistics for English language teaching (ELT). To do so we will introduce key concepts in this area as well as potential applications for pedagogy, including introductory worksheets for teaching collocations in English as a second and/or foreign language (ESL/EFL) contexts with an online corpus.

Key Concepts in Corpus Linguistics and Language Teaching

According to Reppen (2010), a *corpus* is "a large, principled collection of naturally occurring texts (written or spoken) stored electronically" (p. 2). Good reasons to use such corpora (plural) are that they help identify both "linguistic and situational co-occurrence patterns" and offer "a ready resource of natural, authentic, texts for language learning," in addition to removing some of the guessing of native speakers of languages, whose intuitions are "often ill-informed" (p. 4). Thus corpora and DDL are of great value to both native- and non-native-English speaking teachers who question their own intuitions of English language use.

A corpus linguistics approach to teaching and learning is largely *frequency-based* (using a cost-benefit view – see Barker, 2007) with an emphasis on *authentic, real-life examples* that are to be examined *in context.* While the machines of years ago may have caused frustration for most people, corpus linguistics has flour-

ished in the last 20 years with technological advances now providing more user-friendly access to corpora, especially on the Internet. Corpus linguistics and DDL require a corpus, a computer, and concordancing software to analyze the corpus (usually built-in to online corpora sites), plus specific questions to research and answer, and a process for studying the results. We hope to offer examples of the latter two.

While corpus linguistics is used in many linguistic areas (McCarthy & O'Keefe, 2010), language learning and teaching is a primary one, especially for ESL/EFL. The applications of corpora in teaching are myriad, including vocabulary learning and teaching, phraseology, register, English for academic and specific purposes, and materials design. Our examples in the Appendices focus on teaching collocations. Key classroom approaches for DDL with corpora include using word lists, concordance lines for examples of real language use, texts tagged with parts of speech, and examples of register use in language (Reppen, 2010). Our focus in this article is on simple tasks that teachers who are new to corpora and DDL can implement.

A final important concept to note is that DDL requires that teachers and students take on new roles in class. As Gilquin and Granger (2010) clearly state, for teachers it is important to have adequate knowledge about corpora and some of the potential options for using them in the classroom, but DDL also implies something of "a less central role…than in traditional teaching," because the teacher facilitates learning and helps learners arrive at answers to their questions (p. 366). As a result, teachers cede the role of expertise to the corpus, "take risks, and agree to 'let go' and let the student take pride of place in the classroom" (p. 367). Similarly for students, DDL requires some training, as well as guidance in developing questions, determining what resources are available, and the means for understanding and evaluating information that corpora provide.

Approaches and Resources for DDL

Evison (2010) outlines the basics of corpus analysis as "generating frequency lists and concordances" (p. 122). The Key Word in Context (KWIC) approach to concordance analysis enables people to consider a targeted word or phrase to be displayed for easy analysis, centered:

three options we have are a **shopping** centre, a park or entertainment
Bournemouth has got enough **shopping** centres I suppose … The people won't go

know about that, erm, the **shopping** mall. I'm not so sure about the
(Evison, 2010, p. 130)

This example allows us to see that 'shopping' is not just an activity (from the verb 'to shop'), but may also function as an adjective modifying the place where people shop. KWIC examples allow students and teachers to see lexical and grammatical patterns in the various examples offered.

Sinclair (1991) defines a concordance as "a collection of occurrences of a word-form, each in its own textual environment" (p. 32). A concordance example with "up" from the Corpus of Contemporary American English (COCA) is included in Appendix A, shown in Search 4. While KWIC is only one approach to displaying concordance lines, it is likely the main one that most language teachers and learners use. Tribble (2010) discusses concordance searches for words and/or phrases, many of which will involve four steps. However, for more complex words or structures, his full 7-stage process is as follows:

1. *Initiate* a search by looking for patterns in a set of concordance lines (as in the short KWIC example for "shopping" above).
2. *Interpret* the concordance line results by noting patterns (e.g., shopping can be an adjective, so "a shopping centre" or "mall" seems to be a place where one shops).
3. *Consolidate* by looking for further additional patterns in your concordance lines, and report them to others (perhaps ARTICLE + shopping + NOUN).
4. *Report* your results to others by writing out your explicit, testable version of what seems to be happening (e.g., "shopping" precedes the noun it modifies).
5. *Recycle* your results by looking for further information and patterns in this and other contexts in your list of concordance lines (this might evaluate additional findings).
6. *Result* – focus your results for additional study, perhaps incorporating what other authorities (dictionaries, etc.) state that is relevant to your concordance examples.
7. *Repeat* the process above with more data from your corpus, as helpful. (Adapted from Tribble, 2010, pp. 179-182.)

For the "shopping" example, steps 1-4 may suffice. For more complicated words or expressions, one may wish to do the three additional steps where we re-

cycle our findings, focus on the results, and determine whether we need to repeat the process again for further examples and information.

Lee (2010) helpfully introduces key corpora available for teachers to use for concordance or other analyses; see also our starter list of useful online corpora and other resources in Figure 1.

Ten Corpora for ESL/EFL Teaching and Research

British Academic Written English (BAWE) corpus:
http://www2.warwick.ac.uk/fac/soc/al/research/collect/bawe/

British National Corpus (BNC): http://www.natcorp.ox.ac.uk/

☞ See also the Brigham Young University BNC site: http://corpus.byu.edu/bnc/

Collins Wordbanks Online (Bank of English):
http://www.collins.co.uk/page/Wordbanks+Online (requires a subscription)

☞ Corpus of Contemporary American English (COCA): http://corpus.byu.edu/coca/
Academic Vocabulary Lists in the COCA: http://www.academicvocabulary.info/

Corpus of Global Web-Based English (GloWbe): http://corpus2.byu.edu/glowbe/

Corpus of Historical American English (COHA): http://corpus.byu.edu/coha/

☞ Michigan Corpus of Academic Spoken English (MICASE):
http://quod.lib.umich.edu/m/micase/

Open American National Corpus (OANC): http://americannationalcorpus.org/

Santa Barbara Corpus of Spoken American English:
http://www.linguistics.ucsb.edu/research/santa-barbara-corpus

☞ Vienna-Oxford International Corpus of English (VOICE):
http://www.univie.ac.at/voice/index.php (lingua franca English)

Other Useful Corpus-related Internet Resources

Humanising Language Teaching (journal - see esp. its ☞ "Ideas from the Corpora" articles): http://www.hltmag.co.uk/ (geared to EFL, but applies to ESL and other contexts)

Learner Corpus Association: http://www.learnercorpusassociation.org/

☞ Tom Cobb's "Compleat Lexical Tutor": http://www.lextutor.ca/
(many resources, including concordances, tutorials, research, and teachers' pages)

Word Frequency Lists and Dictionary from the COCA: http://www.wordfrequency.info/

(Note: ☞ = particularly recommended.)

Figure 1. Online corpora and other useful resources for ELT and DDL.

Most corpus resources, like those listed above, also allow teachers and students to consider not only English grammar and vocabulary, but also phrases and phraseology. Scott (2010) rightly suggests that computers and corpus software do many things well, such as generating word lists and recognizing patterns, but they cannot analyze data or tell us what it means. That is where teachers who understand

basic corpus linguistic approaches and DDL come in, and can assist their ESL/EFL students in analyzing and understanding corpus results. In our experience, when students conduct their own searches and work with corpus results, this inductive approach makes their ESL/EFL learning more meaningful and memorable.

Implications of Using DDL for ELT

Before moving to pedagogy, let us consider some implications of using DDL in the classroom for English language teaching. While our activities and materials may look similar to traditional approaches, using DDL emphasizes *real examples,* as opposed to fabricated ones we might find in some dictionaries, resource books, or textbooks. One way to ensure this is by using corpus-based textbooks and materials. We can also use tools (like those listed in Figure 1) to search for relevant examples for our classes and to answer our own and our students' questions.

As Gilquin and Granger (2010) indicate, using DDL places the teacher in more of a facilitator role, rather than being the "expert" in class. Overall, we do not perceive this to be a problem, as our goal is to help learners acquire real English language knowledge. Teachers using corpus approaches and DDL can also model and help students to use online and other tools in order to help them learn how to use corpora to learn, by posing questions and finding answers.

We appreciate Sripicharn's (2010) suggestions for preparing learners in class for using corpora by first finding out what they know and providing relevant information on corpora and DDL; second, identifying tasks and related types of corpora to use; and, third, providing access to online corpora or preparing relevant corpora (such as the ones in Figure 1). In addition we may, fourth, introduce corpus analysis tools such as concordancers, as well as possible types of queries with them; and, fifth, help students interpret the results of corpus searches, so they can implement what they learn and know how and when to start again. This cycle, like the one outlined earlier from Tribble (2010), recognizes that as we pose and answer questions with corpus results, repeating the process to arrive at the most useful results may be necessary.

Finally, it is worth noting that although it is possible to use printed concordance lines from corpus research with students, it is now often expected that teachers and students have access to computers during classes, so each teacher will have to determine just how to proceed.

Potential Tasks for the Classroom

Many great books (e.g., Anderson & Corbett, 2009; Bennett, 2010; Campoy, Belles-Fortuno, & Gea-Valor, 2010; Flowerdew, 2012; Reppen, 2010; Shaw, 2011; Sinclair, 2004) introduce, in much greater detail, various ways that teachers can use corpora and English corpus results in their classes. Accordingly, in this article we aim to describe some things that teachers who are not already doing so might begin to incorporate using DDL, in simple, practical ways.

Perhaps the first way that teachers can use corpora for ESL/EFL teaching is in creating and using word frequency lists. In addition to using already established English word lists (e.g., see Lessard-Clouston, 2012/2013, for an overview) and the new Academic Vocabulary List (Gardner & Davies, 2014), for example, with the online Corpus of Contemporary American English (or COCA; Davies, 2008-), teachers can prepare specific lists of English words and then use them either in the preparation of lessons and materials for particular classes or for students to study. Beyond individual words, however, we can also use corpora to help us find collocations and teach lexical chunks (Lessard-Clouston, 2013), as shown in the worksheets included in the Appendices. Like Kathpalia and Ling (2008), we believe collocations can help students develop their proficiency in English, and that using collocations helps learners go beyond individual words and enables them to see that language tends to work in larger phrases or chunks. An excellent resource for much more detail on such approaches to vocabulary teaching using corpora is McCarthy, O'Keefe, and Walsh's (2010) helpful book.

Yet corpora are not only useful for vocabulary teaching. Another way that corpora are valuable is in how they help in examining grammar patterns in English. As alluded to earlier, corpora and concordance lines with real English examples can help both teachers and their students consider specific words and phrases in spoken or written contexts, where they can study preposition use, verbs and complements, and adjective + noun and/or other combinations that will assist English language learners and teachers consider their own speaking and writing. A final task that should be mentioned is that students and teachers may use corpora to compare students' English use (spoken or written vocabulary, grammar, phraseology, etc.) with that of native English speakers (e.g., using the COCA) or other ESL/EFL students (e.g., with the English as a lingua franca Vienna-Oxford International Corpus of English, or VOICE, corpus).

Pedagogy for DDL: The COCA and Collocations

It is not an overstatement to say that online corpus websites are difficult for the uninitiated to navigate. For American English, perhaps the most useful online corpus is the COCA (Davies, 2008-), introduced earlier. As teachers, it has taken us multiple attempts to use the COCA website and much time spent studying tutorials to come to a point where we would feel comfortable doing an impromptu search of the COCA in front of an ESL/EFL class.

We believe that many teachers would say that they think corpora are powerful tools, but that they do not feel ready to teach their students to use them extensively, due to their own limited knowledge. In terms of pedagogy, then, we wanted to create worksheets to help both teachers and students understand the basics of using the COCA online and to feel empowered to use it on their own. In part this is because we believe the COCA is very helpful for teachers and learners to use as a tool for investigating English phraseology and for correcting coded writing errors on their own. Thus, please refer to the Appendices, where our first worksheet introduces the COCA and how to use it to learn which words go together as collocations, and the second one (building on the first) addresses correcting written collocational errors in ESL/EFL.

We chose the COCA because it is a freely available, well-respected website with a very large corpus of over 450 million words of American English. We focused specifically on collocations and collocational errors because learning proper collocations is an area of both interest and difficulty for students. Students may be expected to learn collocations just by taking in a large amount of comprehensible input, and we agree with McCarthy, O'Keefe and Walsh (2010) that concordances can serve as a shortcut to this type of language exposure.

The two worksheets in Appendices A and B are designed to double as lesson plans. They can be used to serve as the very first introductions to using corpus websites. However, while not essential for successfully completing these worksheets, it could be useful to first familiarize students with the skill of reading concordance lines. Teachers can use printouts of concordance lines to do this, perhaps with the four questions outlined by Sripicharn (2010) for observing patterns. Our worksheets do not fully explore the COCA functions, which allow users to compare words, explore synonyms, narrow a focus down to specific genres, and look at words across time. They also do not introduce the very useful search syntax for

searching for a lemma, which collects all parts of speech for a word, plus its inflected and reduced (e.g., n't) forms.

The following worksheets include much explanation to help teachers and students learn to think about search strategies and to navigate the COCA for their own searches. A very self-motivated student could perhaps complete the worksheets on his or her own, but in reality the number of students who would undertake such study would likely be few. Our expectation is that ESL/EFL teachers would go through the worksheets on their own to familiarize themselves with the techniques, and then walk through the scenarios and examples with their students in class, with each student being given worksheet copies to follow along with and work through. Instructors could then assign further searches as homework. They could also give error examples in class and have students discuss (but not execute) the strategies they would use to discover the correction using the COCA. Students could then be asked to try out their proposed methods as homework, with students reporting back in the following class on their findings.

The potential of corpora in the hands of knowledgeable students is great, but the task of helping students learn to use corpora on their own can be daunting. We hope that the worksheets will serve as a useful introduction. Please read through Appendix A for an introduction to using the COCA to help ESL/EFL students learn which words go together, and Appendix B for one approach to helping students learn to use the COCA to correct their written, coded errors.

Final Suggestions for Corpora and DDL Pedagogy

In addition to suggesting that readers explore the COCA using the worksheets provided, we would like to end with some further suggestions for using corpora and DDL pedagogy. Overall, we suggest that teachers feel free to start small, perhaps by using a corpus-based textbook and considering how its examples, vocabulary, etc., differ from other materials they use that are not based on corpus findings. See Lessard-Clouston (2012/2013) for examples of textbooks and other resources focusing specifically on ESL/EFL vocabulary and corpus-based word lists. In addition, we encourage teachers to use all of the corpus linguistics resources that publishers offer related to their textbooks. As teachers are able, we encourage them to explore online resources like the COCA and the Compleat Lexical Tutor in order to further inform their materials selection, understanding of

grammar and vocabulary in class readings, etc. These sites (and many others beyond the short list in Figure 1) can help provide word lists, concordance lines, phrasal combinations, etc., to focus on in specific classes you teach. If you are looking for book-length guides, please consult Bennett (2010), Reppen (2010), or Sinclair (2004). Each is teacher friendly and offers specific steps for particular tasks and Internet sites.

If you have corpus-based resources (textbooks, reference materials, etc.) in class, we suggest you introduce these to your students and gear simple exercises and activities to helping them explore DDL. You might go through the steps outlined by Sripicharn (2010) and Tribble (2010), for example, in terms of using concordances, before attempting to use the worksheets in the Appendices here. We highly encourage you to access whatever resources and tools are available to you, like the online tutorials in the COCA. And don't feel shy about asking colleagues or students who are familiar with the technology to assist you, and help you to learn. Finally, you can refer to the References for good articles and books, as well as Figure 1 for a starter list of online corpora and other web resources that can help with using corpora and implementing DDL in your specific ESL/EFL courses, with their range of levels and skills. Hunston (2010) explains how to use corpora to explore various patterns and Römer (2011) offers a good academic introduction to corpus applications in teaching, while Van Zante and Persiani (2008) provide great examples of using corpora with their ESL classes.

Conclusion

In this article we have outlined practical concepts in corpora and language teaching and introduced basic approaches to and resources for data-driven learning in ELT, including implications for DDL and potential tasks for ESL/EFL classrooms. Whether a refresher for those who know about such options or an initiation into DDL for others, we hope that this overview and the worksheets in the Appendices here will assist those who wish to incorporate online and other corpus resources in their ESL/EFL teaching, especially in the area of collocations.

References

Anderson, W., & Corbett, J. (2009). *Exploring English with online corpora.* New York: Palgrave Macmillan.

Barker, D. (2007). A personalized approach to analyzing 'cost' and 'benefit' in vocabulary selection. *System, 35*, 523-533. doi:10.1016/j.system.2007.09.001

Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers.* Ann Arbor, MI: University of Michigan Press.

Campoy, M. C., Belles-Fortuno, B., & Gea-Valor, M. L. (Eds.). (2010). *Corpus-based approaches to English language teaching.* London: Continuum.

Davies, M. (2008-). *The Corpus of Contemporary American English: 450 million words, 1990-present.* Available online at http://corpus.byu.edu/coca/.

Evison, J. (2010). What are the basics of analyzing a corpus? In A. O'Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 122-135). London: Routledge.

Flowerdew, L. (2012). *Corpora and language education.* New York: Palgrave Macmillan.

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics, 35,* 305-327. doi:10.1093/applin/amt015

Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O'Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 359-370). London: Routledge.

Hunston, S. (2010). How can a corpus be used to explore patterns? In A. O'Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 152-166). London: Routledge.

Kathpalia, S. S., & Ling, K. S. (2008). Is collocation the way to language proficiency? *TESL Reporter, 41*(1), 1-11.

Lee, D. Y. W. (2010). What corpora are available? In A. O'Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 107-121). London: Routledge.

Lessard-Clouston, M. (2012/2013). Word lists for vocabulary learning and teaching. *The CATESOL Journal, 24,* 287-304.

Lessard-Clouston, M. (2013). *Teaching vocabulary.* Alexandria, VA: TESOL International Association.

McCarthy, M., & O'Keefe, A. (2010). Historical perspective: What are corpora and how have they evolved? In A. O'Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 3-13). London: Routledge.

McCarthy, M., O'Keefe, A., & Walsh, S. (2010). *Vocabulary matrix: Understanding, learning, teaching.* Andover, U.K.: Heinle Cengage.

Reppen, R. (2010). *Using corpora in the language classroom.* Cambridge: Cambridge University Press.

Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics, 31,* 249-274. doi:10.1017/S0267190511000055

Scott, M. (2010). What can corpus software do? In A. O'Keefe & M. McCarthy (Eds.), The Routledge handbook of corpus linguistics (pp. 136-151). London: Routledge.

Shaw, E. M. (2011). *Teaching vocabulary through data-driven learning.* Provo, UT: Brigham Young University. [http://corpus.byu.edu/coca/files/Teaching_Vocabulary_Through_DDL.pdf]

Sinclair, J. M. (1991). *Corpus, concordance and collocation.* Oxford: Oxford University Press.

Sinclair, J. M. (Ed.). (2004). *How to use corpora in language teaching.* Amsterdam: John Benjamins.

Sripicharn, P. (2010). How can we prepare learners for using language corpora? In A. O'Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 371-384). London: Routledge.

Tribble, C. (2010). What are concordancers and how are they used? In A. O'Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 167-183). London: Routledge.

Van Zante, J., & Persiani, R. (2008). Using corpus-linguistics findings in the classroom: A rationale with practical applications. *The CATESOL Journal, 20,* 95-109.

About the Authors

Michael Lessard-Clouston (Ph.D., Toronto) is a professor in the Department of Applied Linguistics and TESOL at Biola University in La Mirada, California, where he coordinates the M.A. in applied linguistics and teaches in both the on campus and online M.A. TESOL programs. His research and teaching interests include corpus linguistics, intercultural communication, second language acquisition, and vocabulary learning and teaching.

Tracey Chang (M.A., Biola University) teaches academic ESL at Bethesda University of California and English composition at the Salvation Army College for Officer Training at Crestmont. Her research interests include technology-enhanced language learning, vocabulary learning and teaching, and learner autonomy.

Appendix A
Corpora and Collocations 1
Using the COCA to Learn Which Words Go Together

Introduction: Do you sometimes know a word but aren't quite sure how to use it correctly? It would be helpful to have examples, wouldn't it? A *corpus* is a giant database of the English language full of sentences from books, magazines, and TV. We are going to explore a corpus website called the Corpus of Contemporary American English (or for short, the COCA, http://www.americancorpus.org), where you can look at lots of sentence examples that can help you understand how Americans speak and write in real, everyday English.

You can use the COCA to discover lots of information about language. Today we are going to use it to find out more about words that go together. Words that often go together are called *collocations.* These words seem to be such good "friends" that when you try to put them next to another similar word, it just sounds odd or even wrong to native English speakers.

Here are some examples:

Americans say…	but they don't say….
his **little** sister	his **small** sister
do your homework	**make** your homework
face a **problem**	face a **blessing**
my car is **not working**	my car is **broken**
loyal **to** each other	loyal **for** each other

Let's use the COCA to look at the words *pick up.* This is a special type of grammatical collocation called a phrasal verb. You can say:

(a) "Can you **pick Karen up**?" and (b) "Can you **pick up Karen**?"

It would be logical to think that you can say:

(c) "Can you **pick me up**?" and (d) "*Can you **pick up me**?"

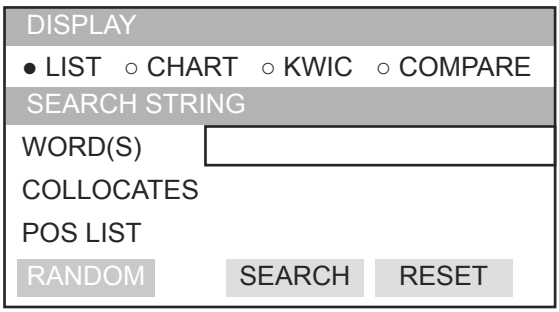
But wait! Did you notice something is wrong? Question (d) is incorrect. You cannot say "*Can you **pick up me**?" It would also be incorrect to say, "Yes, I can **pick up you**." However, it would be correct to say, "Yes, I can **pick up Karen**." Strange, right? There is actually a rule that would help you to know where you can or cannot put the object of the phrasal verb *pick up.* If you had a list of example sentences, you might be able to figure it out on your own. We can use the COCA to do just that. So, the question we want to investigate is, "What is the rule that tells me when I should put the object between *pick* and *up* or when I should place it after *pick up*?" Our strategy will be to look for a list of example sentences containing the phrasal verb *pick up.*

Search 1: LIST + WORD(S)

Question: What is the rule that tells me when I should put the object between *pick* and *up* or when I should place it after *pick up*?"

Strategy: Look for a list of example sentences containing the phrasal verb *pick up.*

1. Go to www.americancorpus.org
2. You will see a search box on the left side that looks like this:



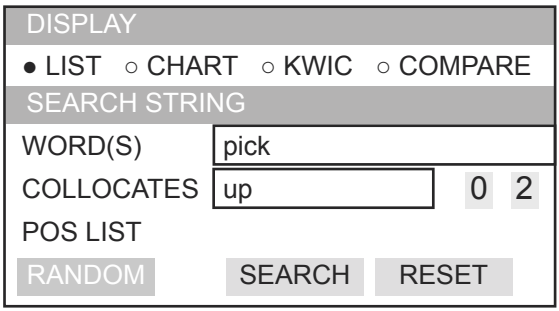
3. Type *pick up* in the WORD box, then click SEARCH.
4. When the results appear in the window, click on PICK UP.
5. You will see a list of sentences (called *concordance lines*) that include ***pick up.*** But these sentences only show *pick* and *up* right next to each other. There are no examples of *pick* + object + *up*. That's because when we typed pick up into the search box, we were asking the corpus to only show us a list of sentences that contain the exact words, *pick up.*

Search 2: LIST + COLLOCATES

We will need to do a new search that includes **pick up** and **pick** + **object** + **up**. We can ask the COCA to show us a list of sentences that contain the word **pick** (Step 2 below) and also contain **up** in the position that is immediately to the right of ***pick,*** or in the position that is two spaces to the right of pick (Steps 3-5).

1. Click RETURN TO SEARCH FORM (top left). Click RESET to start over.
2. Type *pick* in the WORD box.
3. Click on the word COLLOCATES. A box will appear to the right, with two numbers.
4. Type *up* in the COLLOCATES box.
5. Change the two numbers so that it says *0* on the left and *2* on the right. This tells the COCA that you want examples where the word *up* is in the position

to the right of *pick* (pick up) or in the position next to that (pick _____ up). It tells COCA that you don't want examples where *up* appears to the left of *pick* (~~up~~ ~~pick~~). The box will look like this:



6. Click SEARCH.
7. When the results appear in the window, click on UP. (You can also click on the number.)
8. Look at the sentences where ***pick*** and ***up*** are separated by the object (***pick*** + object + ***up***). Write some of the objects here:

_____

9. Which part-of-speech category do most of these objects belong to?
   nouns pronouns verbs adjectives adverbs prepositions
10. Hopefully you noticed that they are mostly pronouns, with some nouns also. Now look at the sentences where pick up is together and the object is to the right.
11. Look at these objects. Which part-of-speech category do most of them belong to?
    nouns pronouns verbs adjectives adverbs prepositions
12. Did you notice that these objects are only nouns? There are no pronouns.
13. So we learned that a noun or pronoun can be placed between *pick* and *up* but only a noun and not a pronoun can be placed to the right of *pick up*.

Search 3: KWIC + COLLOCATES

Sometimes it can take a lot of work to figure out rules about the patterns in language. There is actually a different type of search method we could have used which may have made it easier to find the pattern. We are going to examine *pick up* again, but we will use this other method. This method uses the Key Word in Context (KWIC) display option.

1. Click RETURN TO SEARCH FORM (top left). Do not click RESET.
2. In the DISPLAY options, select KWIC.

3. Click SEARCH.

Wow, we get very colorful results! Each colored word belongs to the same part-of-speech (POS) category. To find out which color represents which POS category, look at the top of the page for a box that says RE-SORT. There is a question mark (?) next to it. Click on the question mark. You can then see what the different colors mean. Today we want to remember that nouns are turquoise blue (bright blue), pronouns are medium blue, verbs are pink, and adverbs are orange. Now click the "back" button on your web browser to return to the results page.

Do you see that all the instances of *pick* are pink (verb) and all the instances of *up* are orange (adverb)? In a phrasal verb, *up* actually belongs to the part-of-speech category called "particles" but the COCA calls it an adverb. Now when we look at the list of examples, we can easily see that there are pronouns (medium blue) and nouns (turquoise) in between *pick* and *up*. We also see that when *pick up* is together, there are only nouns (turquoise) to the right and no pronouns (medium blue). This KWIC display made it much easier to look for patterns, right?

The reason it is easy to see the patterns is because the word to the right of *pick* is in alphabetical order, which is why the repeated words are grouped together. But there are some situations where it would be better if the word to the left of the search word was alphabetical.

Search 4: KWIC + WORD + SORT

Let's say you wanted to find a list of phrasal verbs that have the word *up* as the particle. It would make sense to search for *up* and then look at the verbs on the left, correct?

Question: What are some common phrasal verbs that include *up*?
Strategy: Generate example sentences that include *up*. Use the color coded KWIC format and then look to the left.

1. Click RETURN TO SEARCH FORM (top left). Click RESET to start over.
2. Type *up* in the WORD box.
3. Select KWIC.
4. Click SEARCH.
5. Look at the pink verbs on the left side.

It would be helpful if you could put the words on the left side into order, correct? This is called *sorting*. We want to sort the words to the left of *up*.

Revised strategy: Sort the results so that the word to the left of *up* is in alphabetical order.

To sort the results, you need to understand the sort controller at the top of the page.

It looks like this:

These boxes represent the places where words are located in a sentence. The space in the center represents the search word, also called the *node* word. It makes more sense if you look at the concordance lines in a different way:

Example concordance lines:

village , and then the long walk	up	a steep hill to her house if she didn't
my helmet. Every morning I wake	up	and make my big sister Amy her
tired of Catholic pilgrims showing	up	asking where they might find the
of the lamppost . Caroline looked	up	at the fairyland turret . Did Dr. Cruz
tie worn by a technician who puts	up	evidence displays in court . Two
honest . When Karr failed to show	up	for a court appearance on those
resorts would soon be waking	up	for the season. Much of Swedish

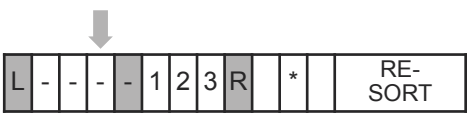
Example concordance lines shown by positions relative to node:

	N-4	N-3	N-2	N-1	Node	N+1	N+2	N+3	N+4
	L	-	-	-	-	1	2	3	R
1	head	and	shoulders	propped	up	a	little	next	to
2	Every	morning	I	wake	up	and	make	my	big
3	Mondrian	,	who	ended	up	asking	where	they	might
4	lamppost	.	Caroline	looked	up	at	the	fairyland	turret
5	Karr	failed	to	show	up	for	a	court	date
6	would	soon	be	waking	up	for	the	time	of
7	and	then	she	looked	up	from	her	book	,

Do you see how the diagram represents the words in the concordance lines (sentences)? Remember that the word in the middle, the search word, is called the *node* word. The numbers **1**, **2**, and **3** at the top show you which position will be sorted first. Right now, the words in the N+1 position are sorted alphabetically. If there is a repeated word in the N+1 position, then N+2 will be alphabetized. You can see an example in lines 5 and 6 above, as this happens when the word **for** is repeated.

Let's go back to the exercise. We want to sort the concordance lines by the word on the left, the N-1 position.

1. Go to the sort controller at the top of the page.
2. Click on the position to the left of the node word.



When you click on it, it will change to say **1**.

You are allowed to select three positions to sort. If you clicked two more spaces, they would change to say **2** and **3**. But for this exercise, we are only interested in sorting by N-1 position on the left.

3. Click RE-SORT.
4. Look at results. Now when you look to the left of *up* and look for pink verbs, it is much easier to quickly find verbs that are often grouped together with *up* and you will have discovered some common phrasal verbs which contain *up*.

Appendix B
Corpora and Collocations 2
Using the COCA to Correct Written Errors

Introduction: You can use the COCA to correct collocational errors in your writing. Your instructor might use the abbreviation **coll** to let you know when there is an error with a collocation in writing. If you don't know the right words that go together, it can be difficult to correct these errors. You can use the COCA to help you find the correct collocations. We'll look at three collocational errors from student essays and practice using the COCA to correct them.

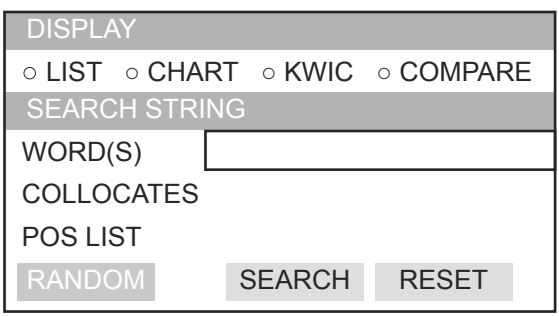
Sentence Error 1

coll

It is very comparable from the beginning of the story.

Question 1: Which word should go after *comparable*?
Strategy: Look for example sentences and look at what words often appear to the right of *comparable.*

1. What type of search would you use to find out this information? Complete the chart below to show which display option you would choose and which word(s) you would type into the search box.



2. Now, go to the COCA website (www.americancorpus.org) and try your search! [If you have it, revisit the Corpora and Collocations 1 worksheet for any steps you would like to review.]
3. After looking at the words which appear to the right of comparable, write down which one you think is appropriate to correct Sentence Error 1.
4. Write the correct sentence here: _____

_____

Sentence Error 2

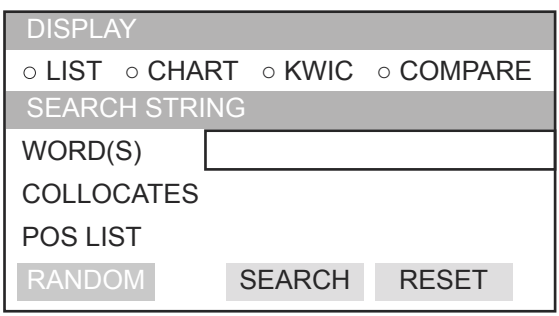
coll

Then he makes a compliment about his father's singing.

This time we aren't sure exactly which word or words are incorrect, but we know that there is a problem with the underlined portion as a collocation. We know that compliment is the key word, so we can start with that.

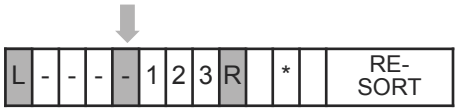
Question 2: How do you use *compliment*?
Strategy: Look for example sentences.

1. What type of search would you use to find out this information? Complete the chart below to show which display option you would choose and which word(s) you would type into the search box.



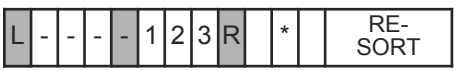
2. Now go to the COCA to try your search.
3. What did you discover? If your search used KWIC display and looked for the word *compliment*, you may have noticed that *compliment* appears as both pink (verb) and turquoise (noun). So we've discovered that *compliment* can be used as both a verb and a noun! What should we do next? If you said that we should use the sorting controller and sort the node position, you are right!

4. Go back to your results and click the node position (center), then click RE-SORT.



Now the nouns (blue) and the verbs (pink) are grouped together. This will make it easier for us to look for patterns.

5. Examine the concordance lines with *compliment* as a noun first. How is *compliment* used as a noun? It might be helpful to sort again. You might notice the collocation "pay a compliment" but you probably won't find anything that says "pay a compliment about _____ ."
6. Examine the concordance lines with *compliment* as a verb. It will be helpful to re-sort the lines again. Sort by the node word position first, then N+1 and N+2. Remember, you simply click on the three spots in the order you want to sort.



How is compliment used as a verb? Do you notice any patterns?

You may have noticed the following patterns:

compliment (someone/a person) on (something that is being admired)
or
compliment (someone/a person) for (something that is being admired)

7. Now correct Error Sentence 2 and write it here: _____

_____

Sentence Error 3

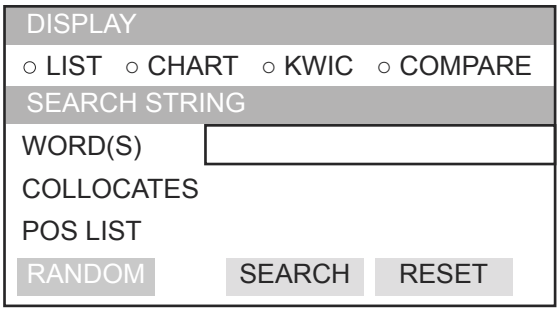
coll

Telling the truth can <u>make people hurt their emotions.</u>

This time we know that *make* and *emotions* are incorrect words in this collocation, but we know that *hurt* is correct.

Question 3A: What is the correct collocation for *hurt*?
Strategy: Look for words that go with the key word *hurt.*

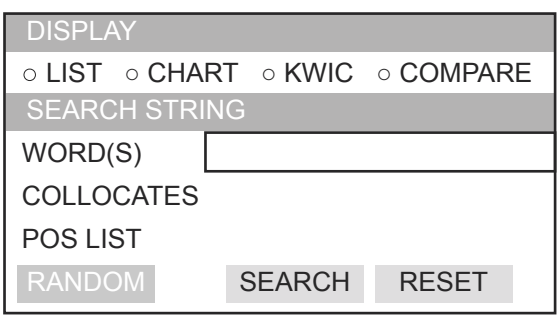
1. What type of search would you use to find out this information? Complete the chart below to show which display option you would choose and which word(s) you would type into the search box.



2. Now go to the COCA to try your search. Can you find a word that would have a meaning similar to *emotions*?
3. A basic LIST search of the word *hurt* would show the word *feelings* as the top search result. Now click on the word FEELINGS to see how to use it correctly.
4. Answer the following questions:
   Does it seem ok to say "hurt their feelings"? Y N
   Does it seem ok to say "make people hurt their feelings"? Y N

Question 3B: What are the correct collocations for hurt with feelings?
Strategy: Look for more examples of "hurt their feelings" to find out how to use it.

5. What type of search would you use to find out this information? Complete the chart below to show which display option you would choose and which word(s) you would type into the search box.



6. Now go to the COCA to try your search.
7. Can you find any patterns for how *hurt their feelings* is used? Write down any patterns you have discovered: _____

   _____
8. Now correct Sentence Error 3 and write it here: _____

   _____