Title

Using Corpus Linguistics and Language Analysis Tools for ESLVocabulary Instruction: Two Case Studies

Permalink

Journal

ISSN

Authors

Rodriguez, Emmanuel
Csomay, Eniko

Publication Date

DOI

Copyright Information

Peer reviewed

Emmanuel Rodriguez
MiraCosta College

Eniko Csomay
San Diego State University

Using Corpus Linguistics and Language Analysis Tools for ESL Vocabulary Instruction: Two case studies

Choosing relevant vocabulary to teach in ESL composition courses can be a difficult task for instructors. They often rely on intuition rather than empirical methods for vocabulary selection. This paper presents two case studies where corpora were used either to select vocabulary items for instruction or to scaffold student learning of vocabulary through exposure to corpora. Corpora and language analysis tools were used to complement vocabulary selection and instruction. This vocabulary selection process complements perceptual approaches as it provides empirical measures for the selection and a data driven approach to learning as instructors teach and students learn contextually relevant words.

Keywords: Corpus-informed teaching materials, multilingual learners, English for Academic Purposes

Introduction

Vocabulary is an important component of language instruction in ESL courses. To teach relevant vocabulary, instructors need to make important decisions to identify the words to be taught throughout their courses. Although many instructors rely on vocabulary words provided in textbooks or choose to use vocabulary lists such as the Academic Word List (AWL; Coxhead, 2000), the target vocabulary in these materials is often not what students need to meet their learning goals. As a beginner instructor, choosing relevant vocabulary words can be a difficult task since the first impulse is to rely on intuitions rather than other measures (e.g., information gained through frequency counts) when selecting these vocabulary items. That is, instructors are more likely to choose words that they think are more relevant and pertinent to the students' proficiency level and content of the class, perhaps leaving out important information and insights about the words themselves.

In this paper, we present a case study from two ESL writing courses through which we showcase how corpora and language analysis tools can be applied to complement the intuitive nature of decisions for choosing and teaching relevant vocabulary words, and by doing so, we also provide a replicable framework.

Literature Review

Vocabulary learning is one of the most important foundations for second language development at any proficiency level. Learning the content words of a discipline is crucially important for academic success at all levels of education but particularly in higher education. In college-level English for Academic Purposes courses, as a college-level writing course would be, reading and writing skills work together as students read and talk about different texts on the same topic, after which, they are typically tasked to write about the topic while utilizing the content words from the readings. Although students need to have a vocabulary base of 8000-9000 word families in order to understand a written text (Nation, 2006), it is impossible for instructors to teach all the vocabulary words that their students need to know (Nisbet, 2010) during the period of a course. Therefore, the instructor's selection of vocabulary should prioritize the words that will help students understand the readings and the topic discussed in class.

Additionally, instructors often rely on native-speaker intuition to make the appropriate selection of vocabulary but oftentimes ESL instructors are non-native speakers themselves where intuition may not be a reliable source. In these instances, having a reliable framework available to aid the decision of choosing vocabulary words is crucial. Biber and Reppen (2002) discuss the importance of using authentic materials that reflect language use. Following their school of thought, corpus linguistic (CL) techniques can provide empirically-based results which can then complement how instructors and curriculum designers select vocabulary words. In addition, they can provide students with the opportunity to explore authentic language use. More specifically, corpora and corpus tools allow instructors to rely on high frequency vocabulary words that their students will need to complete the assigned language tasks.

A Data-driven Approach to Vocabulary Instruction

Vocabulary word selection can be a difficult and overwhelming task for instructors. As mentioned before, many instructors rely on intuition rather than empirical methods when selecting the words that are helpful to students. Friginal (2018) presents data-driven learning (DDL) as an effective, learner-centered approach to vocabulary teaching and learning. Although it is a great way to promote vocabulary learning, it is important to note that, as Hunston (2002) points out, DDL does not teach a language pattern; instead, it helps instructors and students visualize authentic uses of the language to draw conclusions about patterns of language use based on the samples they retrieve from corpora. Hence, DDL is an approach where students interact with real language compiled in a corpus as they identify patterns of language use. It functions as a tool for students to explore the language and its patterns beyond a limited scope (Friginal, 2018) that instruction may provide. Some linguists have argued that DDL is not pedagogically appropriate because, if not introduced properly, it could feel monotonous and it, in itself, lacks induction of meaning (Nolen, 2020). To make it right, DDL requires instructors and curriculum designers to spend considerable time and effort in learning how to utilize DDL and CL properly. These aspects (e.g., time and effort put into making the activities right, especially the first time around) may be concerning to instructors but the benefits outweigh the concerns.

Over the years, linguists have found benefits of using DDL in language learning classrooms. DDL vocabulary instruction offers learners "relevant and meaningful" (Friginal, 2018) language data along with frequency distributions and language patterns. For instance, saliency and meaningful language interactions are two important characteristics of DDL vocabulary instruction, allowing students to explore language patterns beyond a "prescribed" approach. According to Friginal (2018), "[C]orrectness and accuracy in using language, however these are defined, are clearly important constructs in CL, but instead of focusing on or prioritizing prescribed (i.e., 'correct') forms, actual frequencies of use, not intuitions, alongside a full attention to and consideration of contexts, are established in the forefront" (p. 39). CL offers a DDL approach to vocabulary instruction where language instructors use empirical measures for word selection.

Not only instructors can use corpora as a resource and CL techniques as a tool for word selection, but students may also explore how vocabulary words are used in authentic texts. There are different approaches to how students learn with CL techniques. One approach is explicit learning where students have the intention to analyze and study words, their meanings, and parts of speech variations. A different approach is incidental learning where students pick up vocabulary words from their contexts while noticing their features (Zimmerman, 2014). Friginal (2018) shares different corpus-based activities where students can find language patterns and identify important vocabulary words depending on the subject of study. In one of the activities, students are tasked to analyze authentic texts and identify academic and technical words for the topic discussed. Adapted from Friginal (2018), in the activity below (see Figure 1), the instructor pre-determines the vocabulary words that students would look for. After the words are chosen, students are given instructions that guide them through the activity. Students access the Corpus of Contemporary American English (COCA) (Davies 2008-) and type one of the predetermined vocabulary words, "*impact*," in the search bar. The following is an example of what these instructions look like.

Figure 1

Activity worksheet adapted from Friginal's (2018, pp. 218-221) activity description

---

Step 1. Consider the following excerpt from the article "Survival Factors"

The initial impact with the seawall occurred at 1127:50 … Some flight attendants stated that the first impact was followed by a sensation of lifting off again.

What part of speech is "impact" in this context?

Step 2. In COCA, type, "impact" in the search bar and choose the part of speech that matches the context described earlier.

Possible definitions of "impact":

    (1) The striking of one body against another
    (2) A forceful consequence
    (3) Influencing strongly
    (4) The violent interaction of individuals or groups entering into combat

Which definition do you think fits the context of this excerpt?

Step 3. Which synonyms could you use for "impact" in the context of this excerpt? Choose two:

Contact   Collision
Shock   Bang
Crash   Brunt
Blow

Step 4. Examine the collocates for *impact*. Click on a verb collocate and take a look at the concordance entries. Do you see any patterns? Create two original sentences that contain the collocate and *impact*, with at least one sentence using the topic discussed in our class.

Step 5. Choose another collocate that is an <u>adjective</u>. Take a look at the concordance entries. Do you see any patterns? Particularly, notice prepositions. Create two original sentences that contain the collocate and *impact*, with at least one sentence using the topic discussed in our class.

In the activity presented above, students have the opportunity to interact with authentic materials, notice patterns of language use, and utilize the sentences they found in the corpus as a mentor text to write their original sentences. Besides, this activity helps students notice how the vocabulary item is used in a specific context. As shown in Step 1 of the activity, students are asked to look at the excerpt provided and think about the part of speech (POS) category of the vocabulary item. Instead of teaching students explicitly the POS and asking them to memorize it, this activity requires them to notice the POS from the context including the sentence structure. In Step 2, students are tasked to choose a meaning from the list of definitions provided that matches the POS in the corpus. In Step 3, students are shown a list of synonyms and are tasked to choose the words that could be synonyms in the context discussed. In Step 4, the corpus allows students to notice language patterns and find collocations. Additionally, students have the opportunity to use the results from the corpus to write original sentences using the vocabulary words in focus. Instructors should make sure that they provide feedback to students' work in order to reinforce learning. It is important to note that before this activity, students should be taught what collocations and POS are, so this activity can work

effectively. In the last step, students are asked to find collocates that are in a different part of speech category. This is one of the activities that Friginal (2018) provides to illustrate how effective and beneficial CL is in ESL vocabulary instruction. In the following section, we discuss different CL approaches to vocabulary instruction.

CL Approaches to Vocabulary Instruction
When choosing vocabulary words for instruction, language instructors should consider using authentic language and perhaps could consider associated frequency counts. Liu and Jiang (2009) discuss a lexicogrammatical approach to teaching collocations. In their study, they suggest that vocabulary and grammar learning take place simultaneously. In support of their claim, their research showed that students benefited from analyzing the grammar structures that were used along with the vocabulary words they were supposed to learn. In this approach, as students learn new vocabulary items (e.g., afraid, pay, etc.), they also learn the collocations that frequently co-occur with these words that at times end up being grammatical collocates (e.g., afraid of, pay for, etc.). This approach results in a better understanding of how to use these collocates accurately when students are producing a text.

Another example is when a word has different meanings depending on the context. In this case, language learners can benefit from analyzing how the word is used in different contexts and situations. Hunston (2002) supports the importance of teaching these different meanings and exemplifies this phenomenon by using the word, "maintain" which can be used to 1) not allow to weaken something (*she has maintained her relationship*), 2) say something strongly (*maintain a claim*), and 3) to keep something at a certain level (*the prices have been maintained low*). Instead of teaching students a vocabulary word in isolation, students benefit from learning how these words can be used in different contexts along with the collocations that frequently co-occur with the chosen words.

<center>

The Current Study

</center>

This paper presents two case studies where corpora were used either to select vocabulary items for instruction (case study #1) or to scaffold student learning of vocabulary through exposure to corpora (case study #2). The motivation for the case studies presented next is based on the following principles:

- Providing a data driven approach to vocabulary instruction and vocabulary learning
- Finding an empirically-driven framework to vocabulary selection for ESL instruction
- Utilizing language analysis tools and corpora to inform vocabulary selection, instruction, and learning

In the following section, the procedures we followed when selecting vocabulary words from the articles students read in the courses taught are explained.

The educational context
The two courses discussed in these case studies are the first composition courses that first-year undergraduate English learners and international students at a university level have to complete to prepare for higher level courses. These courses serve as an introduction to writing at a college level for first-year English learners and prepare students to understand college level readings, rhetorical contexts, and writer's arguments. In these courses, students integrate ideas from multiple sources in formal written papers. English learners and international students need to complete three major writing courses in this program. Only the first two courses in the program are discussed in this paper.

Course #1 offers first-year English learners and international students an introduction to college level reading and writing. In this course, the major writing assignments are as follows: a response essay, a case study, and an argumentative essay. This course is also divided into three units. In unit 1, students read an article and write a response essay. In units 2 and 3, students read a total of eight articles on the same topic (the cognitive and social effects of bilingualism) and write the case study and the argumentative essay while connecting ideas and synthesizing information from these

articles. Similarly, in course #2, students complete three major writing assignments: a narrative essay, an exploratory synthesis essay, and an argumentative essay. To complete these assignments, students read several academic texts and integrate them into their writing assignments. The class is divided into three units, so each unit focuses on each of the major writing assignments. In unit 1, students focus on reading three narrative essays in order to write their own narrative essay. In units 2 and 3, students read a total of five articles on the same topic (the role of college education) and synthesize information from these articles to complete the exploratory synthesis and the argumentative essay.

For this case study, we chose units 2 and 3 for course #1 and course #2. The reason for choosing these units for each of the courses is because the topic is the same for all the articles in the two units for each course. Students may struggle with the vocabulary and language complexity of the topic. Therefore, using a language analysis tool allows us to identify high frequency words that students need to know to better understand the topic discussed in each course.

The context for the two case studies

For this case study, one of the researchers taught the classes in discussion. He compiled two different corpora, one made of the articles that students in course #1 and course #2 read for units 2 and 3. The topic discussed in course #1 is the cognitive and social effects of bilingualism, and the topic discussed in course #2 is the role of college education. Table 1 summarizes general information about both.

Table 1
The corpus of reading materials for two courses

Course #1	
Article	**No. of words**
A1	1375
A2	727
A3	2360
A4	1759
A5	791
TOTAL	7012
Course #2	
Article	**No. of words**
A1	1239
A2	1163
A3	756
A4	1024
A5	737
A6	822
A7	862
A8	2696
TOTAL	9299

Case Studies

Case-study #1

One of the researchers and instructor of the course used *AntWordProfiler*, a freely available software (Anthony, 2022)*,* which allows the user to find the most frequent words throughout the articles. However, this program does not categorize the words into parts of speech groups or into content versus function word groups. Hence, as you see in Figure 2, the instructor found that many of these words were determiners and prepositions in high frequency and only a few were content words which are of interest in this short study.

Figure 2
High frequency words across the readings for course #1

	type	headword	list_id	freq
6	the	the	3	310
2	of	of	3	231
10	in	in	3	200
9	and	and	3	195
24	to	to	3	192
49	a	a	3	174
78	that	that	3	137
3	language	language	3	92
46	is	is	3	79
84	they	they	3	55
205	are	are	3	53
206	not	not	3	53
83	their	their	3	52
81	or	or	3	51
134	on	on	3	51

Since the most frequently used words in Figure 1 do not inform the instruction of important content words, the Academic Word List (AWL) was used (Coxhead, 2000) to filter the search to the most frequent and academic words appearing in the corpus. The AWL is used because it consists of 570-word families and 3110 individual words that are commonly used in academia. After filtering the search, the results indicate that the top five most frequent academic words in the five articles are *acquire*, *academic*, *foundation*, *benefit*, and *investing* as shown in Table 2. The topic discussed in the articles was bilingualism and its advantages, so these words seem to be connected to the topic. However, there are some words, for instance, *minority* and *majority*, that frequently appear in the texts but may not be explicitly related to the topic of bilingualism. In situations where words do not connect directly with the topic discussed, the instructor needs to make decisions based on the importance of the words in relation to the topic discussed in the course.

Table 2
Filtered high frequency words for course #1

Word	Frequency
acquire	3
academic	3
foundation	3
benefit	3
investing	2
majority	2
minority	2
research	2
adulthood	1
attainment	1

After limiting the search to words that also appear in the AWL, the instructor used *AntConc* (Anthony, 2022), which is a corpus analysis tool for concordance and text analysis, to analyze these words and to identify collocations that frequently co-occur with these words. He found that the words *acquire*, *academic*, and *foundation* do not occur with collocations in the texts. However, the word *benefit* frequently co-occurred with the word *from* as in *'benefit from'*, as shown in Table 3. This finding shows that the collocation 'benefit from' can be taught together, instead of simply teaching the word 'benefit' in isolation. As mentioned before, corpus linguistics allows students to identify these patterns and replicate them in their own writing.

Table 3
Concordance lines for the word "benefit"

LEFT CONTEXT	Collocate	RIGHT CONTEXT
research shows that both minority- and majority-language children **benefit**	**from**	dual language education. For example, in a district-wide

US needs to embrace multilingual education — our children will **benefit**	**from**	it.
language to native English speakers so they too can **benefit**	**from**	the cognitive, neurological, economic, and cultural . that investing in

In cases where the corpus has a limited number of collocations to draw from, a larger corpus representative of language use can be accessed. There are multiple advantages of using a larger corpus. First, using a large corpus, such as the Corpus of Contemporary American English (COCA), allows learners to access a wealth of information. Second, such a corpus allows learners to filter results according to their interests. For instance, if they are interested in spoken language use, they can filter the results by spoken English. Similarly, if learners are interested in academic written English, they can filter results to fill their needs and interests. Third, a corpus like COCA provides concordance lines as well as graphs, so learners can better understand differences in the contexts where the target vocabulary words are used. On the other hand, one of the disadvantages is that using such a large corpus does not allow for specific topics and subfields. It may be the case that the words discussed in the articles and readings do not actually display the collocations shown in the corpus. Although it is beneficial to teach the words with collocations that appear in the larger corpus, these collocations may not match the collocations used in the articles.

To illustrate the process discussed previously, the instructor used the Corpus of Contemporary American English (COCA) to find collocates of the word, *"acquire"*. As explained before, the word *"acquire"* did not display relevant collocates in *AntConc* with the limited corpus. He searched for the word *"acquire"* in COCA, using the collocates tab which allows users to view frequent collocates of the target vocabulary word and filtered the search to academic words. Figure 3 shows the most frequent collocates of *"acquire"* as shown in COCA.

Figure 3
COCA frequency count for right collocate for the word acquire

★	WORDS	FREQ	ALL	%	MI	
★	KNOWLEDGE	225	49966	0.45	5.23	
★	STUDENTS	222	227824	0.10	3.03	
★	SKILLS	205	39818	0.51	5.43	
★	INFORMATION	116	78663	0.15	3.62	

As noted in these results, the words that frequently co-occur with acquire are mostly nouns. Most of these nouns (e.g. knowledge, students, etc.) seem to connect with the topic of bilingualism. These words can be taught within the context. However, there are other contexts where the word *"acquire"* may not relate to the context. For example, one of the results in the corpus was *"acquire land"* which in this case, does not connect to the topic of bilingualism and it has a slightly different, more general meaning (to obtain). This does not mean that it should not be taught, but students could benefit from learning the term within the specific context discussed in the class and can learn how words can have multiple meanings as they enter specific disciplinary domains versus having a general meaning in other, more general domains. As mentioned before, this is one of the advantages of using corpus linguistics. Students are able to explore language patterns and identify how the target vocabulary word is used in different contexts.

Case-study #2

After the instructor of this course tested the procedure outlined above for the articles for course #1, he used the same procedure in the second corpus (course #2) to find out whether or not he would be able to find collocations in both corpora. One of the reasons for testing the same procedure in a different corpus is because two of the articles in course #1 are written by the same author, which could affect the reliability of the study. He used the articles for course #2 in the second exploration. There are more articles, and they are spread more evenly in word quantity. Table 1 shows more details about the articles for course #2.

Without filtering the search to the AWL list, he found similar results where the most frequent words were closed categories as shown in Figure 4.

Figure 4
High frequency words across the readings for course #2

	type	headword	list_id	freq
15	the	the	3	419
2	to	to	3	319
24	and	and	3	298
9	of	of	3	234
58	a	a	3	211
6	in	in	3	180
70	that	that	3	156
21	for	for	3	122
52	college	college	3	109
64	students	students	3	87
39	are	are	3	85
62	is	is	3	85
32	it	it	3	71
68	as	as	3	71
147	skills	skills	3	68

Once he identified high frequency words, he filtered the search results to words that matched the AWL and found that 9.9% of the words in all texts also appeared in that list. Table 4 shows the top 10 words with the highest frequency.

Table 4

Filtered high frequency words for course #2

Word	Frequency
Liberal	14
Outcomes	11
Job	9
Global	6
community	6
data	5
somewhat	4
technology	3
respondents	3

He used *AntConc* (Anthony, 2022) to analyze these words and to find out if there are collocations in the texts students read. Figure 5 shows the collocates of *'liberal'* as shown in *AntConc*. From these collocates, the most frequent and relevant to the topic discussed would be "*arts*" as in *'liberal arts.'* This finding is interestingly important because the term *'liberal arts'* is more specific to the topic discussed as opposed to the word *'liberal'* in isolation. Although teaching students the word *'liberal'* in isolation could be beneficial, to understand the text, students need to have a foundation of the term *'liberal arts'* as a whole, instead of two separate pieces of information.

Figure 5

Collocates of the word "liberal" in AntConc

	Collocate	Rank	FreqLR	FreqL	FreqR	Range	Likelihood	Effect
1	arts	1	16	1	15	3	91.289	5.402
2	emphases	2	2	1	1	1	14.234	6.402
3	skills	3	8	2	6	1	13.131	2.314

Additionally, the word *'outcomes'* also shows collocates that frequently co-occur as shown in Figure 6. In this finding, the words *rank*, *essential*, and *value*, do not really connect to the topic discussed in this unit (the purpose of higher education). However, the collocation *'learning outcomes'* has a deeper connection to the topic discussed. When he analyzed the collocation more in depth, most of the concordance lines show *'student learning outcomes'* together as one concept. The word *'outcomes'* in isolation is important to teach, but for students to understand the text within its context, it is highly beneficial to teach the collocation.

Figure 6
Collocates of the word "outcomes" in AntConc

	Collocate	Rank	FreqLR	FreqL	FreqR	Range	Likelihood	Effect
1	rank	1	2	0	2	1	14.034	6.402
2	essential	2	3	2	1	1	13.796	4.665
3	learning	3	4	4	0	1	13.035	3.647
4	value	4	3	2	1	1	11.843	4.179

Scaffolding vocabulary learning through exposure to corpora
Using the previous list of words (Table 6) found for the texts that students in course #2 read, we demonstrate a vocabulary learning activity where students can interact with corpus to notice language patterns that frequently occur with the target vocabulary words.

Activity
Assuming that the instructor of the class had already selected the vocabulary words from Figure 6, the purpose of this activity is to activate students' schemata about the words and the content discussed. To enhance the effectiveness of this activity, it is highly recommended that the instructors use an activity worksheet as the one shown next to ensure students follow directions. Pairing students and assigning roles are equally important and can be beneficial since students would be able to help each other.

The first part of this activity (see Figure 7 in the appendix for the activity's description) asks students to read an excerpt from one of the articles. The excerpt contains the word, "outcomes", which is one of the target words to be learned. Students are asked to guess the part of speech category of the word. After students answer the first question, they should enter COCA in a browser and use the "Word" tab that allows detailed information about the word. Next, students can notice that words are considerably frequent in academic genres as opposed to other genres. Once students compare frequencies, they choose a definition provided in COCA. Some words may have more than one definition, and other words could have only a single definition. After students learn the meaning of the word, the next step is to help them notice the collocates that frequently co-occur with the target word. Students use the "Collocates tab" in COCA to find the most frequent collocates with the target word. Last, students use this part of the corpus to find authentic language use and write original sentences using the concordance lines as mentor texts.

Discussion and Implications
The case studies presented in this paper aimed to provide ESL instructors a replicable framework to use corpus linguistics and language analysis tools to complement vocabulary word selection and to teach vocabulary. The framework provided is not a definitive way to choose and teach vocabulary words, but its purpose is to complement vocabulary instruction by providing authentic language patterns, ease of access, and student autonomy.

These short case studies illustrate a framework that can be used for different contexts. However, as with all frameworks, there are limitations. One of the limitations here is that the corpus compiled for both courses is not sufficient to do detailed analyses of language patterns. In other words, we need a larger corpus to make assumptions about vocabulary words used in different contexts. As a result, some of the words provided did not have a connection to the topic discussed. Another limitation is the amount of work that instructors would need to perform for every class in order to choose relevant vocabulary items. Even though this can be a limitation, for those who teach the same content every semester, the framework presented could help since vocabulary selection process would only be completed once and would be improved every semester. Lastly, the framework provided is dependent on technology. Therefore, if the corpus and the language analysis

tools become unavailable for a period of time, the instructor will need to find a different option to supplement the technology used in the framework.

In conclusion, the framework provided can be highly useful for specific contexts. For instance, for classes in English for Specific Purposes, students deal with the same topical knowledge frequently. This framework provides instructors with reliable and empirical data to inform instruction as well as students with authentic language use and enables them to explore language use in different contexts.

Authors

Emmanuel Rodriguez is a college professor at Southwestern College, San Diego City College, and Miracosta College. He is also associate coordinator at the English Center in San Diego City College where he develops professional development for embedded tutors in English and ESL courses. His research interests include educational technology for language learning, corpus linguistics, and ESL reading and writing.

Eniko Csomay is a Professor at San Diego State University and a corpus linguist with a particular interest in language use at the university. She published articles in highly ranked international journals (Applied Linguistics, Journal of English for Academic Purposes, etc.) and edited many volumes such as, The Routledge Handbook of Corpora and English Language Teaching and Learning (2022). She was awarded multiple fellowships (Fulbright, English Language Specialist) and worked with ESL/EFL teachers abroad (Hungary, Georgia, Mexico, Montenegro, Morocco, Singapore, Taiwan).

References

Anthony, L. (2022). AntConc (Version 4.0.10) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Anthony, L. (2022). AntWordProfiler (Version 2.0.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar instruction? *Studies in Second Language Acquisition, 24*(2), 199-208. https://doi.org/10.1017/S0272263102002048

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238.

Davies, M. (2008-). *The Corpus of Contemporary American English (COCA)*. Available online at https://www.english-corpora.org/coca/.

Friginal, E. (2018). *Corpus linguistics for English teachers: new tools, online resources, and classroom activities (1st ed.).* Routledge. https://doi.org/10.4324/9781315649054

Hunston, S. (2002). *Corpora in applied linguistics.* Cambridge University Press. https://doi.org/10.1017/CBO9781139524773

Liu, D., & Jiang, P. (2009). Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal*, *93*, 61-78.

Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, *63(1)*, 59-81. https://doi.org/10.3138/cmlr.63.1.59

Nisbet, D. L. (2010). Vocabulary instruction for second language readers. *Journal of Adult Education 39*(1), 10-15.

Nolen, C. (2020). Examining task transferability in task-based language teaching: A multi-case study. Ph.D. Dissertation, Georgia State University. https://doi.org/10.57709/16118716

Zimmerman, K. (2014). A literature review of the challenges & best practices for English language learners. *National Forum of Multicultural Issues*, *11*(1), 1-7.

Figure 7
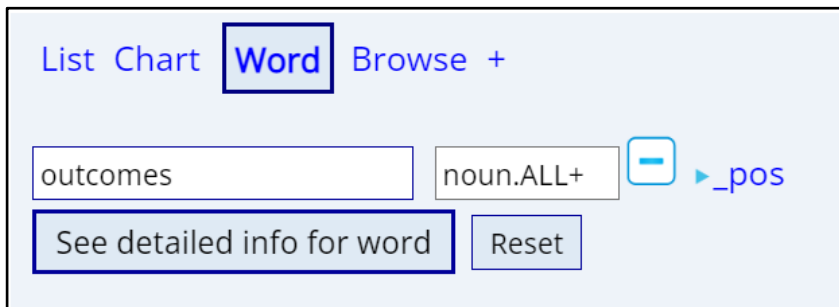*Activity worksheet: Pre-reading activity*

1.- Read the following excerpt from the text, "What Employers Want?" and answer the question below the excerpt.

Excerpt from the text:

"Younger employers -- those under 40 -- place a higher value on civic-related learning **outcomes** and experiences than do employers over 50."

What is the part of speech of the word, "outcomes"? _____

2.- Use COCA and select the "Word" tab. Type in the word, "outcomes" with the POS you selected in question 1.

The search results will give you important information about the word (definitions, frequencies, topics, collocates, synonyms, clusters, and concordance lines).
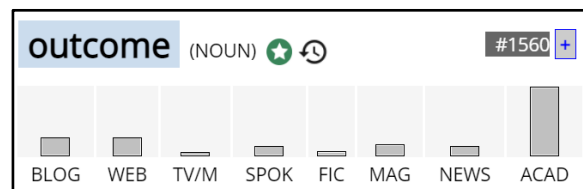
3. Answer the following questions

   a)   List the top 3 genres where this word is more frequently used according to this graph.

Genre 1: _____
Genre 2: _____
Genre 3: _____

   b)   Choose one of the definitions provided and write it in the space below.

---

c) From the list of collocates, notice that the word, "learning" is present as well as in the text's excerpt in question 1. This seems to be the collocation, "learning outcomes."

Go back to the search bar and click in the "Collocates" tab. Type the word, "outcomes" and make sure that the numbers 1 at the bottom are selected to both sides of the word.

List Chart Word Browse **Collocates** Compare KWIC -

| outcomes | Word/phrase | noun.ALL+ | ⊟ ►_pos |
| * | Collocates [POS] |

| + | 4 | 3 | 2 | **1** | 0 | | 0 | **1** | 2 | 3 | 4 | + |

Find collocates | Reset

You will see the most frequent collocates that appear with the word, "outcomes." Click on the collocate, "learning" and notice all the sentences that appear.

HELP	ⓘ	★	WORDS	FREQ	ALL	%	MI	
1	❶	★	HEALTH	1373	300737	0.46	6.34	
2	❶	★	STUDENT	897	147171	0.61	6.75	
3	❶	★	LEARNING	868	125607	0.69	6.93	
4	❶	★	POSITIVE	709	92398	0.77	7.09	
5	❶	★	POSSIBLE	441	213970	0.21	5.19	
6	❶	★	EDUCATIONAL	407	49653	0.82	7.18	

In the space below, write 2 sentences where the collocate is used in a similar way to how it is used in the article's excerpt.

Sentence 1: _____

Sentence 2: _____

d) Find 2 collocates that are nouns and appear frequently with the word "outcomes." Write the collocations below and write an original sentence. Take a look at the concordance lines if you need an example (Some ideas at the bottom of this worksheet).

e) Find 2 other collocates that are adjectives and appear frequently with the word "outcomes." Write the collocations below and write your original sentence. Take a look at the concordance lines if you need an example (Some ideas at the bottom of this worksheet).

COLLOCATES | OUTCOME | NOUN | Advanced options | 🏠 | Collocates Clusters Topics Texts KWIC 🕑 ⬇

+ NOUN		NEW WORD	?	+ ADJ		NEW WORD	?	+ VERB		NEW WORD	?	+ ADV		NEW WORD	?
2364	3.19	student		1670	5.22	positive		1029	4.48	improve		368	4.42	regardless	
1851	3.66	health		1043	3.32	possible		959	4.33	affect		228	7.38	eg	
1459	4.99	measure		976	2.28	different		871	5.16	predict		219	2.08	et	
1390	4.19	election		870	2.85	better		724	3.68	determine		165	5.75	ie	
1178	2.73	study		736	4.55	negative		685	4.95	influence		164	2.86	significantly	