HOW TO USE CORPORA IN LANGUAGE TEACHING FOR VOCABULARY ENHANCEMENT

Nihan ERDEMİR

Sabahattin YEŞİLÇINAR

1. Introduction

Being a crucial part of learning a foreign or second language, vocabulary is crucial for learners in acquiring a language. To become efficient users of the target language, learners should have a good vocabulary repertoire. In other words, an extensive vocabulary size enables learners to communicate in the target language fluently and accurately because vocabulary size indicates language ability (Alderson, 2005). Conversely, those who wish to achieve academic language proficiency need to know about academic success (Nation & Coxhead, 2021). Research has also underlined the correlation between vocabulary knowledge and effective speaking (e.g., Uchihara & Saito, 2019), writing (e.g., Lee et al., 2021), reading (e.g., Schmitt et al., 2011), listening (e.g., Wang & Treffers-Daller, 2017), and discourse and pragmatics (e.g., Szudarski, 2018). Since vocabulary knowledge enables language use and influences four skills, it is beneficial to investigate possible ways to enhance learners' vocabulary repertoire and effectively promote vocabulary development.

Non-native speakers have gaps in their linguistic knowledge, making it difficult to communicate in written or verbal form. They notice these gaps when they are unsure of a particular word choice (Mueller & Jacobsen, 2016). In this situation, they usually refer to dictionaries. Although dictionaries can be practical for lexical choices, "they often provide inadequate information about certain facets of language, such as how words typically combine (i.e., collocations)" (Mueller & Jacobsen, 2016, p. 4). Possible alternatives to dictionaries can be the use of language corpora as a resource in language teaching and learning. Corpora and concordancing software provide authentic input that develops learners' mastery of correct word use (Liu & Lei, 2018) and expand their analytical abilities (Thurstun & Candlin, 1998). They also help learners revise their writing and error-correction. Empirical studies have also reported the positive results of concordance-based vocabulary learning compared to traditional vocabulary learning (e.g., dictionaries) (Cobb, 1999). Integrating corpora into target language vocabulary learning or teaching has various advantages, such as enhancing metalinguistic knowledge and promoting learner autonomy (Yoon, 2008), increasing learners' confidence (Yoon & Hirvela, 2004), developing research skills, and enabling learners to revise their writing more effectively (Mueller & Jacobsen, 2015), drawing attention to the connection between linguistic expressions and context (Aston, 2001), and noticing the gap (Swain, 1995). For this purpose, researchers have focused on using corpora in L2 vocabulary learning based on these benefits, and this sudden interest towards corpus linguistics has occurred by means of free and well-organized corpora (e.g., Corpus of Contemporary American English (COCA, Davies, 2008) though first use of concordances in the classroom dates back to 1969 (McEnergy & Wilson, 1997).

2. Literature Review

2.1. Vocabulary and language learning

The role of vocabulary knowledge in the target language is undeniable because limited vocabulary hinders successful communication. Individuals must use words to communicate their thoughts and ideas in any language. According to Wilkins (1972), little can be transferred without grammar, but nothing can be transferred without vocabulary. Given the importance of vocabulary learning, "lexical knowledge is central to communicative competence and the acquisition of a second language" (Schmitt, 2000, p. 55). However, individuals' native language and target language vocabulary growth are different. For example, in the native language, children acquire most words through repeated encounters in context. However, the primary exposure to the target language occurs mainly in the classroom, and children use their mother tongue for communication outside the classroom. Therefore, it is necessary to implement an intervention in the vocabulary learning process in the classroom so as to develop learners' target language vocabulary (Hart & Risley, 1995). Otherwise, learners will have limited target language input, causing a decrease in their motivation. That is, it is of great importance to plan how vocabulary will be learned, thereby optimizing learning.

Vocabulary learning does not only refer to the mere definition and form of a word. Indeed, it is a process that includes the association and selection of different vocabularies to reflect an idea to be conveyed (Schmitt et al., 2011). However, learners with insufficient vocabulary will face a dilemma in learning a target language, especially in writing and reading while they are using the language (Ahsanuddin et al., 2022). Given the effect of vocabulary mastery on reading comprehension, Schmitt et al. (2011) suggest that learners should know 98% of the target vocabulary to comprehend a text. In other words, learners' vocabulary knowledge predicts their comprehension of a text (Hu & Nation, 2000) and their writing, listening, and speaking skills. In a similar vein, Nouri and Zerhouni (2016) consider vocabulary mastery a crucial part of reading comprehension. As to writing, Hinkel (2011) and Kiliç (2019) focus on the impact of vocabulary capacity on writing skills. Finally, examining the relationship between second language vocabulary knowledge and second language reading/listening comprehension, Zhang and Zhang (2022) report high correlations between vocabulary knowledge and listening skills. That is, vocabulary knowledge consists of several aspects of language learning.

2.2. Corpora use in vocabulary learning

Although various methods have been utilized for vocabulary enhancement in language teaching and learning, corpora have grown in prominence over the past four decades thanks to technological advancement. According to Cobb and Boulton (2015), the effects of the corpus have been of three main types. The first is upstream use, which "lies in improved descriptions of language varieties and features that inform aspects of the language to be taught" (Boulton, 2015, p. 478). The second type "makes corpora and tools for analyzing them available to the teacher" (ibid.). From the teacher's point of view, corpora can be helpful in determining what

to teach. Teachers using the corpora for this purpose do not benefit from large modern corpora like the British National Corpus (BNC: Oxford, 1995) and the Corpus of Contemporary American English (COCA: Davies, 2009). Instead, they use small corpora like Brown (Kucera & Francis, 1979). Finally, the third puts them directly into the learner's hands. The learners' making queries in corpora includes techniques that are essentially akin to corpus linguists' activities: "Like a researcher, the learner has to form preliminary hypotheses based on intuition or scanty evidence; those hypotheses then have to be tested and rejected or refined against further evidence, and finally integrated within an overall model" (Johns, 1988, p. 14).

The literature also provides empirical evidence on the advantage of applying corpora in vocabulary teaching. For example, after having investigated whether teaching vocabulary through corpus-based activities is more effective than teaching vocabulary with textbooks or dictionaries, Paker and Ergül-Özcan (2017) report that using corpus-based vocabulary materials is significantly more effective compared to using textbooks or dictionaries. Moreover, comparing the effectiveness of using corpus examples to dictionary definitions, Frankenberg-Garcia (2014) suggests using the multiple corpus examples because they help learners understand new vocabularies and use them appropriately at a syntactic level. Another research underlines the effectiveness of using data-driven learning in teaching vocabulary (Boulton & Cobb, 2017). The researchers advocate that through concordance lines learners can observe vocabularies in an authentic rich context and discover their meanings in consideration of available contextual clues. To exemplify, Daskalovska (2015) compared the traditional learning of collocations (through matching, fill-in gaps and multiple-choice exercises) with the corpus-based learning (through concordancers) and noted that the experimental group became more successful in the test.

Moreover, previous research on the use of corpora in language classrooms provide pedagogical foundation for authenticity, learner autonomy and error correction. To begin with, a corpus is a collection of naturally occurring language samples of a particular linguistic item, so it is singled out for bringing authenticity into language classrooms (Gilquin & Granger, 2010). This approach is called *data-driven learning* (DDL) which favors the use of corpus tools for pedagogical purposes. Second advantage of this approach is that learners are involved in using condordancing software to investigate both form and meaning in a given context. Therefore, learners' development of autonomous learning strategies is noted as a result of 'discovery learning' with their active involvement (Bernardini, 2002; Pérez-Paredes et al., 2004). Last, corpus allows learners to find out their errors or "interlanguage features (misuse, overuse and underuse)" by comparing their writing with the corpus data and improve their writing skills (Gilquin & Granger, 2010, p. 359). In addition to native speaker corpora, learner corpora, which is called *learning-driven data* (LDD) by Seidlhofer (2002), might help to realize these problematic features of a learner language. For instance, investigating how learners addressed to native speaker corpora and learner corpora in classroom activities, Cotos (2014) suggests the intehration of LDD into classroom activities in order to benefit more efficiently from DLL.

To elaborate, teachers might refer to various aspects of vocabulary by using corpora. It is summarized below how the corpus data allow us to explore the lexicon:

1 Lexis: the word frequency of the general lexicon (e.g., *high frequency items*) and word formation (e.g., *high frequency prefix*)

2 Phraseology: collocation and patterning (e.g., *take refuge*), fixed expressions and idioms (e.g., *of course*)

3 Meaning: context and meaning (e.g., *take something seriously* or *take heed*), polysemy (e.g., *race*), metaphor, connotation and ideology (e.g., '*refuge from'* connotates to undesirable situation)

4 Sets: lexical sets (e.g., topic-specific words in the business texts such as 'fund, share, market'), synonyms (e.g., *seek asylum* or *take refuge*), antonyms (e.g., *unhappy and sad*, or *not happy*)

5 Lexis in spoken language: phraseology (e.g., *take care*), meaning and usage (e.g., *you know*) (Moon, 2010, pp. 197-211).

Considering that the studies on corpus use on vocabulary learning suggest important findings on enhancing vocabulary capacity and reinforcing the four skills, foreign language teachers are suggested to develop and adapt the materials. Despite the large number of publications on corpora use for language learning, there is still limited research that would focus on learners' and teachers' classroom practices in using corpora. Therefore, this chapter will provide several resources and suggest sample activities and materials for corpus use in teaching vocabulary in language classrooms at differing levels of proficiency. For the teachers who have limited knowledge of corpora and thus low tendency to use corpora in the classroom, in-service teacher training is suggested (Mukherjee, 2000). According to Ma et al. (2021) teachers might need *corpus-based language pedagogy* (CBLP) in order to use the corpus tools for pedagogical purposes.

2.3. Guidelines and resources for teachers to develop and practice corpus-based materials

Learners might need to become familiarized with a new approach. Therefore, presentation, practice and production stages could be followed. To begin with the corpora introduction, general information about the basic functions of a corpus and corpus terminology can be provided to students; second, corpus tools can be introduced, and students are allowed to experience in using the corpus tools; last, corpus-based activities can be used to practice in class (Sripicharn, 2010).

There are several freely accessible resources prepared by the experts with a limited number of queries per day. The commonly used ones are listed as below:

English Corpora

- AntConc
- British National Corpus (BNC)
- Compleat Lexical Tutor
- Corpus of Contemporary English (COCA)
- Sketch Engine for Language Learners (SKELL)
- Word and Phrase Info

L2 English Corpora

- Corpus of Academic Learner English (CALE)
- English as a Lingua Franca in Academic Settings (ELFA)
- International Corpus of Learner English (ICLE)
- Louvain International Database of Spoken English Interlanguage (LINDSEI)
- Vienna-Oxford International Corpus of English (VOICE)
- Trinity Lancaster Corpus (L2 spoken English)

3. Sample Activities for Corpus Use in Language Classrooms

Based on the abovementioned literature review, in this section, sample activities are presented for teachers who aim to integrate corpus into their classrooms for different ages, levels of proficiency and purposes. The activities are ranged in accordance with their objectives: corpus use for *understanding meaning and use of words*, *error correction* and *understanding genre*. Teachers may consult to each activity in accordance with their context, needs and interest.

3.1. Understanding meaning and use of words

In this activity, young learners refer to a corpus in order to understand the meaning and use of a word or phrase. A corpus-based website, Sketch Engine for Language Learners (SKELL), is used. It is user-friendly with its simple interface, so the students can easily use. It offers authentic samples and allows to investigate the contextual uses of words under three main functions: *examples*, *word sketch* and *similar words*.

Materials: Computer lab, internet access, a worksheet per student

Level: A2

Age: 11-12

Language skills: reading and writing

Duration: 30 minutes

Purpose: to identify contextual uses of words, to write examples for each word

Procedures:

1 Hand out a text and a worksheet to each student.
2 Ask students to read the text and discuss comprehension questions in groups.
3 Give students time to underline unknown words.
4 Have students open SKELL at https://skell.sketchengine.eu/#home?lang=en.
5 Explain that SKELL can provide up to 40 example sentences per search.
6 Walk the students through a simple search in SKELL.
   a Type the word 'space' to be explored in the search box and browse.
   b Select 'examples' section to see the uses in sentences (see Figure 1).

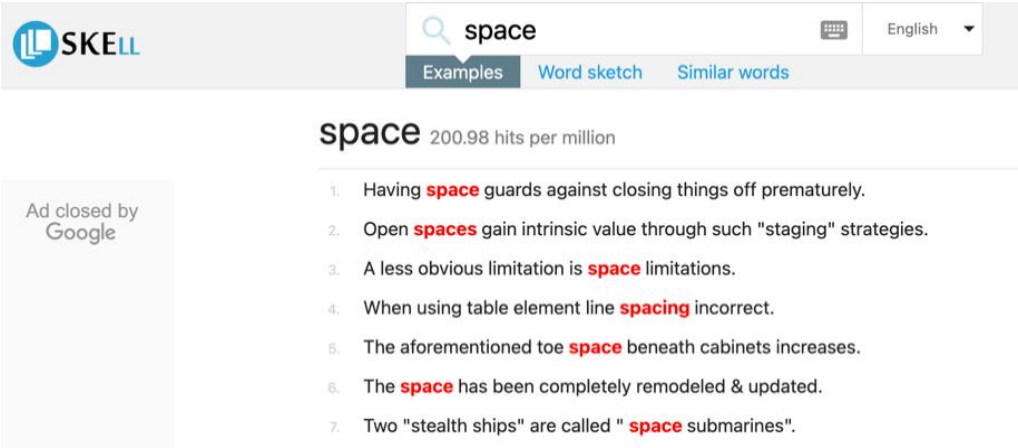


Figure 1. Search results for space in SKELL

   c Scroll down the page, and indicate them to notice many instances as a verb and noun, singular and plural, with and without article.
   d Have students click on 'word sketch' section and analyze the collocations of the word space in accordance with the frequency (see Figure 2).

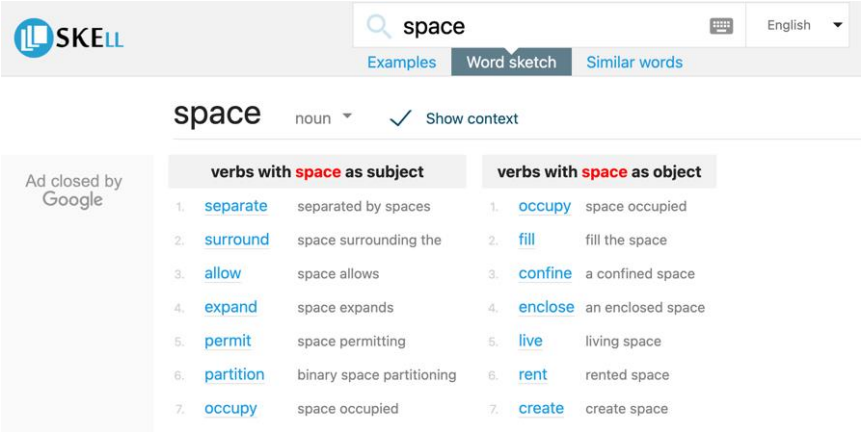


Figure 2. Search results for space in word sketch section

e Have students compare the collocations used in subject and object.

f Ask students to select 'similar words' section to enhance their vocabulary capacity and analyze the synonyms of the word space (see Figure 3).

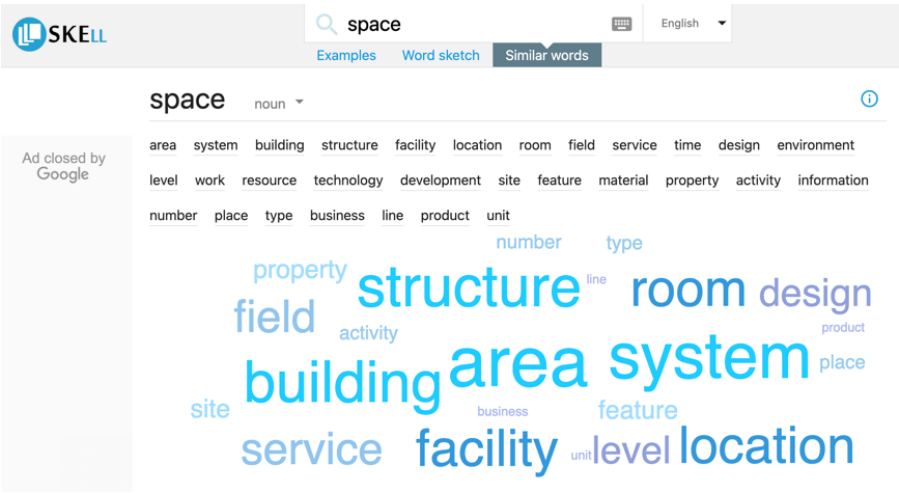


Figure 3. Search results for space in similar words section

g Remind students that they cannot replace each word because of the meaning in context.

7 Ask five minutes for group work and complete the worksheet which can be prepared by the results in SKELL.

8 Allow students to further search in SKELL by themselves to answer.

a For example, ask them to place the adjectives in accordance with the meanings in context.

For example: large / big

A …… space for major events will be located on the fifth floor.

President Obama has already made the ….. space exploration announcement.

3.2. Error correction

This activity requires students to refer to both a general and learner corpus. After writing their draft, students can first browse the native speaker corpora to understand the rules and patterns, and then they can compare and contrast their writing with the learner corpora.

Materials: Computer lab, internet access

Level: B1

Age: 15-16

Language skills: reading and writing

Duration: 30-min

Purpose: to identify contextual uses of words, to compare and contrast the higher frequency and L2 tendency about the uses

Procedures:

1. Ask each student to write a 200-word paragraph on a given topic.
2. Have students to focus on one linguistic aspect in the paragraphs, for example, *moreover*. In accordance with students' common errors, various linking words might be asked.
3. Ask students to analyze how they used in their draft.
4. Pair them to discuss how each used *moreover* in a sentence and compare with peers' draft
5. Indicate their errors with frequency, grammar patterns and meanings.
6. Introduce them first to a learner corpus, ICLE
   a. Ask students to search for *moreover* (see Figure 4).

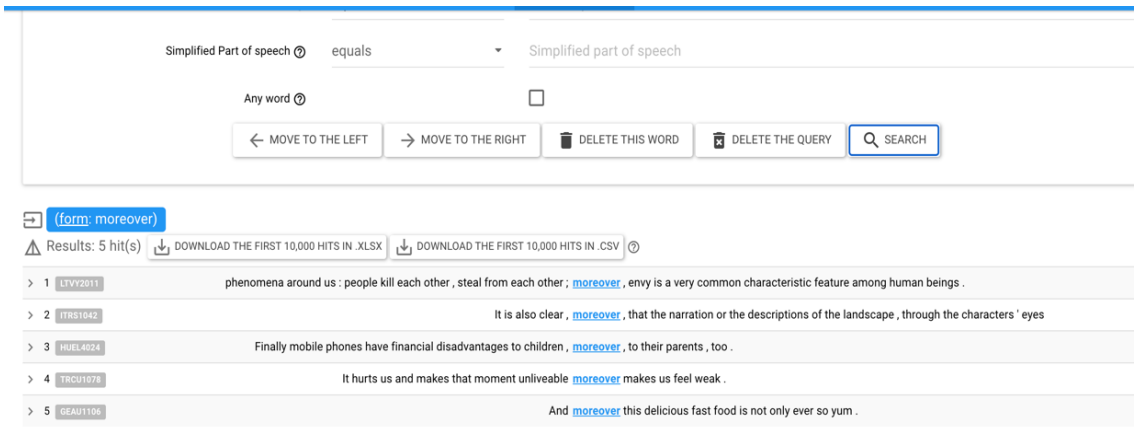


Figure 4. Concordance lines for moreover in ICLEv3

   b. Indicate similar errors by learners. For example, the 4th sentence can be given as an example for an incorrect use in ICLE-Tr which includes examples from Turkish L2 speakers of English.
   c. Ask students to click on the word to check it in the extended context (see Figure 5).

Concordance details

Extended context

Is it a Remedy or an End . Before touching on this concept I would like to emphasize that it is really difficult to decide and it is subjective . We should respect to everyone 's thoughts . Because , no one could understand (if he does not come across with this situation) the person 's mood who have to decide to do that . So my thoughts about this concept depends on only my logic not my experiences . PAIN is such a fenomenon that no one can stand for along time . It hurts us and makes that moment unliveable moreover makes us feel weak . Let 's think about that moment that we cut or burned our finger . What do we feel ? At that moment we just think how to stop that pain , we can not think something else . It is quite simple example from our daily life . Now , let 's think about a person who has an uncureable ilness and pain . He has to stand this pain until he dies . His ilness is such an ilness that it has no remedy , no cure and he lives like a plant by the help of the machines .

CLAWS lemma and POS analysis

form	lemma	part of speech	simplified part of speech
and	and	CC	CCO
makes	make	VVZ	Vlex
that	that	DD1	DET
moment	moment	NN1	N
unliveable	unliveable	JJ	ADJ
moreover	moreover	RR	ADV
makes	make	VVZ	Vlex
us	we	PPIO2	Pronpers
feel	feel	VVI	Vlex
weak	weak	JJ	ADJ
.	PUNC	PUNC	PUNC

Variables

Native language	Turkish
Country	Turkey
Gender	Female
Age	21
Other foreign language(s)	German , NA , NA
Language(s) at home	Turkish , NA , NA
Years of English at school	6 or more
Years of English at university	4
Months in English-speaking country	0
Institution	Turkey - University of Çukurova
File name	TRCU1078
Title	Euthanasia
Length in words	851
Type	Argumentative
Conditions	No timing
Reference tools	No
Examination	No
Date (yyyy-mm-dd)	2003-01-15

Figure 5. Extended context of moreover in ICLEv3

7 Ask students to compare and contrast its use in a native speaker corpus.
8 Invite the class to visit COCA at https://www.english-corpora.org/coca/
9 Ask students to type *moreover* in the search string and hit the 'Find matching strings' button (see Figure 6).

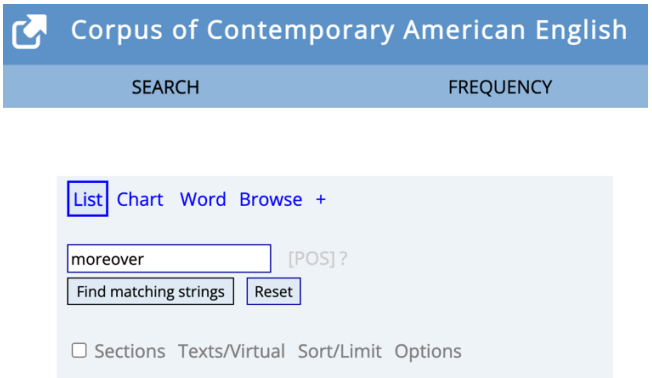


Figure 6. Search for moreover in COCA

10 Ask students to report on the number of occurrences found in COCA.
11 Ask students to click on the underlined word and retrieve all concordance lines (See Figure 7).

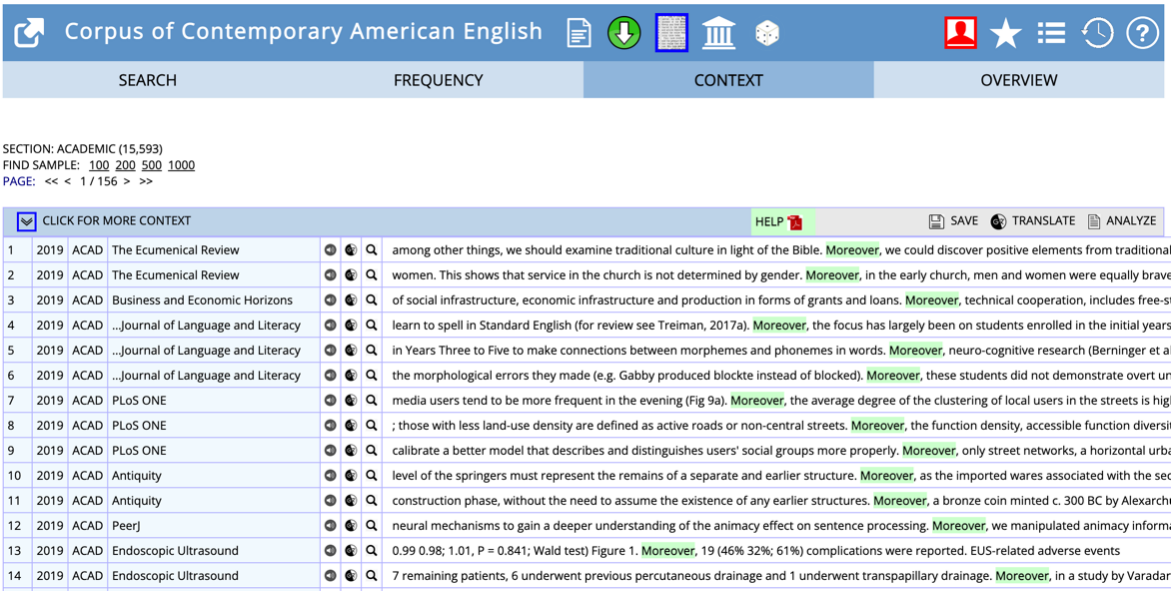


Figure 7: Concordance lines for moreover in COCA

12 Pair students and allow them to compare and contrast their results in terms of frequency, grammar patterns and meanings.
   a Ask students to report their observations
   For example: In ICLEv3, *moreover* is often used in the middle of sentences with two preceding and following commas, while in COCA *moreover* is often used at the start of sentences with one following comma.
13 To further research, ask students to click on '+' and 'Keyword in Context (KWIC)' buttons, type *moreover* and hit the 'Find matching strings' button (see Figure 8).

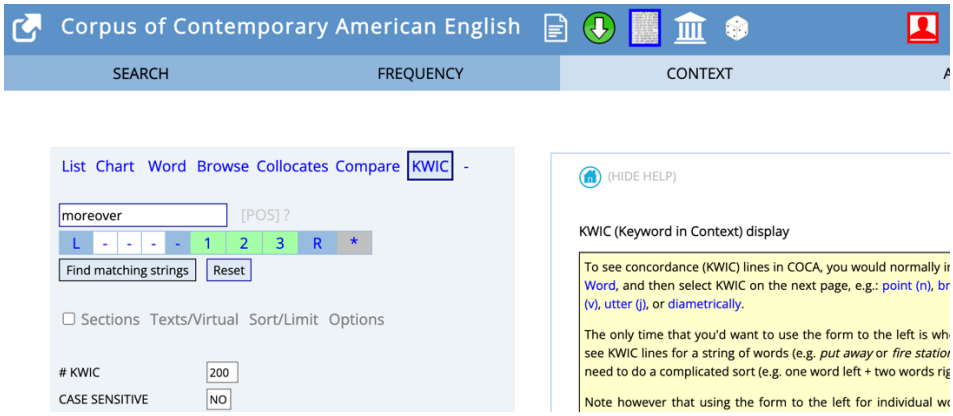


Figure 8. KWIC search for moreover in COCA

14 Have students examine the KWIC results for their search and explain that different parts of speech are indicated in colors (e.g., pink for verbs, yellow for prepositions and conjunctions) (see Figure 9).

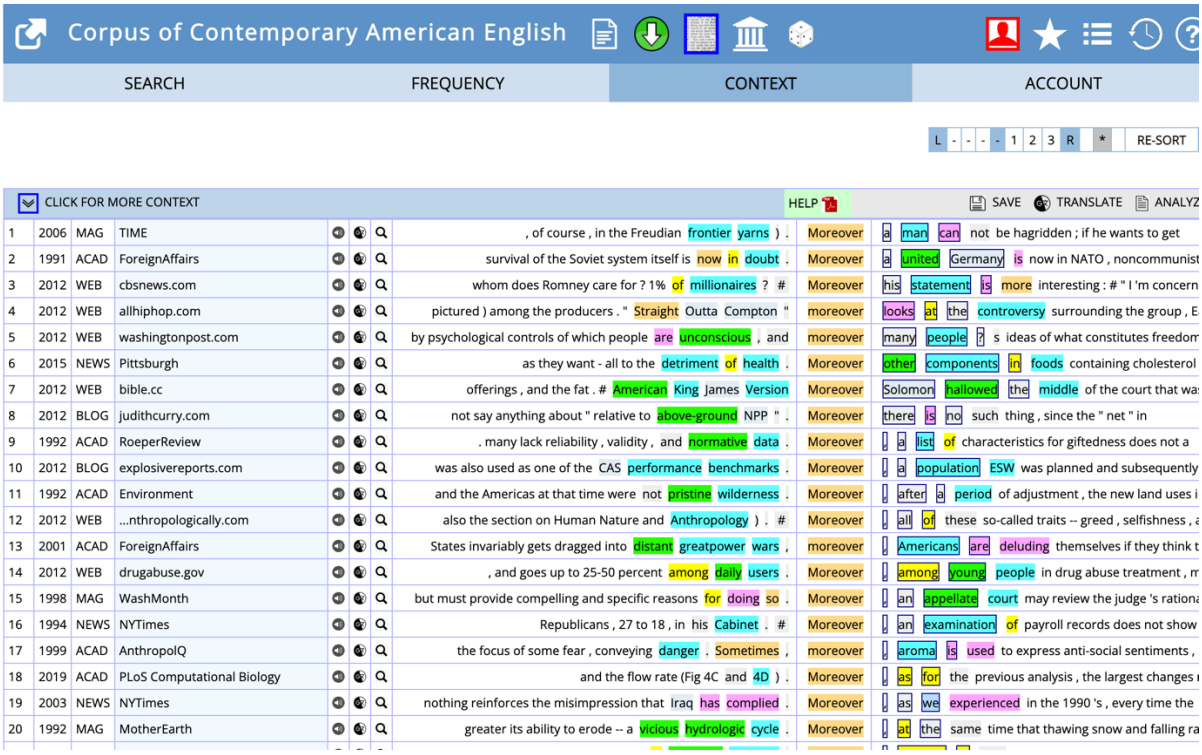


Figure 9. KWIC results for *moreover* in COCA

15 Allow students to discover the uses in various contexts.

3.3. Understanding genre

In this activity, learners are asked to differentiate the word choice which is specific to a text type such as spoken (e.g., *interviews*) and written (e.g., *academic papers*). The collocations are further analyzed to understand their uses in context.

Materials: Computer lab, internet access, a reading text and worksheet per student

Level: B1/B2

Age: 18-19

Language skills: reading and writing

Duration: 30-min

Purpose: to identify contextual uses of words, to raise awareness of word choice in spoken and academic contexts.

Procedures:

1 Ask students to read formal and informal texts.
2 Allow them to discuss informal versus formal language and how written academic discourse differs from the spoken text. To exemplify;

a *grammatical intricacy*: short sentences in spoken discourse / longer and complex sentences in written discourse.

b *lexical density*: higher use of function words (pronouns, prepositions, and articles) in spoken discourse / higher use of content words (nouns, verbs, adjectives, and adverbs) in written discourse.

c *nominalization:* less use of longer noun groups in spoken discourse / more use in written discourse.

3 Have students focus on verbs and verb + noun collocations

4 Ask students to underline verbs used in a spoken discourse and to find their synonyms in an academic discourse.

5 Invite the class to visit the BNC at https://www.english-corpora.org/bnc/

6 Ask students to choose 'List' section and type *give* and *provide* in the BNC, respectively.

7 Ask students to click on 'Sections' and select 'Spoken' for the first box and 'Academic' for the second box (see Figure 10).

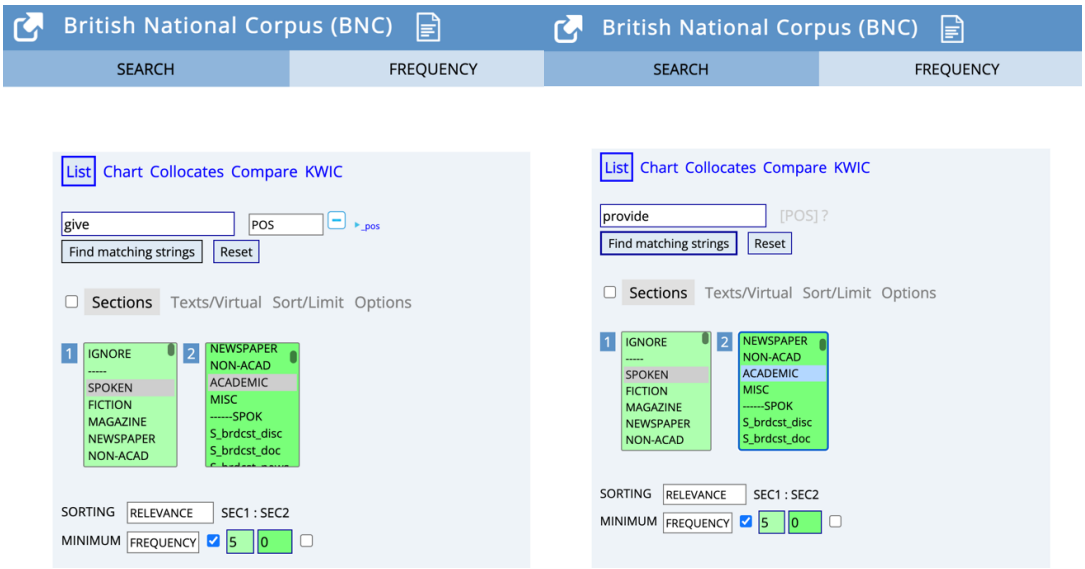
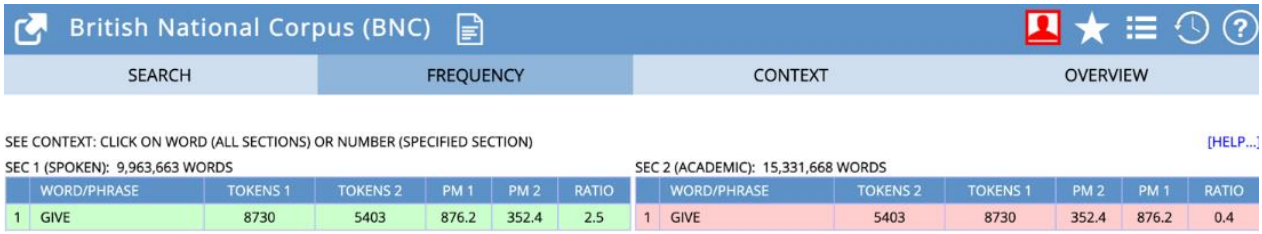


Figure 10. Search for give and provide in spoken and academic contexts of the BNC

8 Ask students to hit the 'Find matching strings' button in order to compare and contrast their occurrences in different discourses.

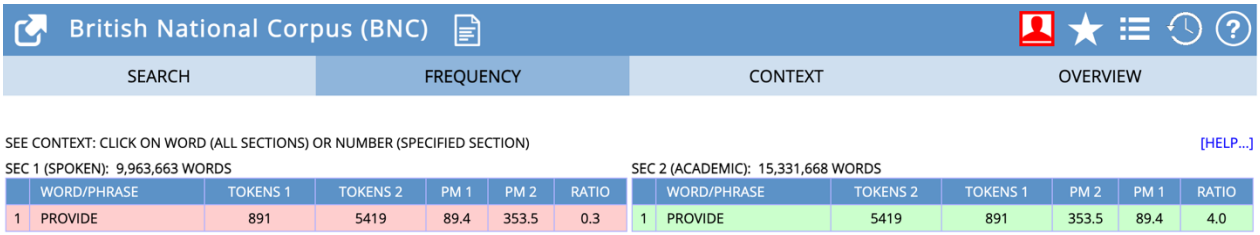
9 On the results page, indicate how give is often used in spoken discourse as shown in Figure 11, while provide is used at a significant level in academic discourse (see Figure 12).



	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1	GIVE	8730	5403	876.2	352.4	2.5

	WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	GIVE	5403	8730	352.4	876.2	0.4

Figure 11. Distribution of give in spoken and academic contexts



	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1	PROVIDE	891	5419	89.4	353.5	0.3

	WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	PROVIDE	5419	891	353.5	89.4	4.0

Figure 12. Distribution of provide in spoken and academic contexts

10 On the results page, indicate how give is often used in spoken discourse as shown in Figure 11, while provide is used at a significant level in academic discourse (see Figure 12).

11 Ask students to click on the number of occurrences in 'Token 1' (see Figure 13), and 'Token 2' (see Figure 14).

British National Corpus (BNC)

SEARCH | FREQUENCY | CONTEXT | OVERVIEW

FIND SAMPLE: 100 200
PAGE: << < 1 / 4 > >>

CLICK FOR MORE CONTEXT | HELP | SAVE | TRANSLATE | ANALYZE

1	JS8	S_meeting	Liberal resolutions that we can get (unclear) it is important (unclear) at the sorry to give evidence of the alternative viewpoint and, and that is the case in, in
2	JJT	S_courtroom	, unquote. (pause) By the tear-- by the time (pause) each of them came to give evidence in this case it was clear that each had brought forward in their own
3	JJU	S_courtroom	who is er the defendant's expert. Erm, both erm experts are to give evidence before (unclear) and my Lord included in this bundle with their reports are the
4	CH1	W_newsp_tabloid	# THE LIMIT # RICK SKY # OVER THE LIMIT # YOU needed balls to give evidence in Roman times. Under ancient Roman law only men with two testicles were
5	CH1	W_newsp_tabloid	Roman times. Under ancient Roman law only men with two testicles were allowed to give evidence in court. Eunuchs and slaves were deemed unreliable by judges. # M
6	CH2	W_newsp_tabloid	to court as libel witness # STEPHEN WHITE # PRINCESS Diana may be forced to give evidence in court about her friendship with dashing Gulf War hero James Hewitt. Sh
7	CH2	W_newsp_tabloid	is unlikely that polo-playing Major Hewitt would want to embarrass the princess by making her give evidence in front of the world's press. Earlier, the paper's money-sp
8	CH3	W_newsp_tabloid	Pakistani paceman, also played in the games but is not expected to have to give evidence at the inquiry. # KENNY BLAST # ALEC JOHNSON # BLACKBURN boss Kenny
9	CH5	W_newsp_tabloid	will hand myself in as soon as they free Boler or I am required to give evidence at Boler's appeal.' That's why I'm on the run
10	ABP	W_ac_polit_law_edu	allow (at that time), one of the parties to an action to give evidence; but it is a procedure, and the only procedure, which is
11	ABP	W_ac_polit_law_edu	and wife are competent, and within narrower limits that they are compellable, to give evidence against one another, in criminal proceedings. Under the Law Reform (Hu
12	ACJ	W_ac_polit_law_edu	that the prosecution need not rely on members of the public to come forward and give evidence, which there is often a reluctance to do. But one effect of
13	CN5	W_ac_polit_law_edu	' pass' if they reach certain levels of competence: they are expected to give evidence of having reached a particular level of performance regardless of how they stand ir
14	EDL	W_ac_polit_law_edu	to try; # (3) # that the witnesses to be examined may give evidence material to the issue; # (4) # that there is good
15	EDL	W_ac_polit_law_edu	value. The defendants contested the claim, and the relevant Russian officials were to give evidence on their behalf. The plaintiff's case rested solely on his own evidence

Figure 13. Search results for provide in a spoken discourse

British National Corpus (BNC)

SEARCH | FREQUENCY | CONTEXT | OVERVIEW

FIND SAMPLE: 100
PAGE: << < 1 / 2 > >>

CLICK FOR MORE CONTEXT | HELP | SAVE | TRANSLATE | ANALYZE

1	HF0	S_brdcast_documentary	Wolves and and Stoke. (SP:PS2WM) Wolves in fact have done more than most to provide evidence that the game is ridding itself of violence. The football autho
2	A1V	W_newsp_brdsht_nat_report	right of parents to terminate the life of a brain-dead daughter. The cases should provide evidence on the conservative course of a court heavily influenced by a
3	K97	W_newsp_other_report	Hafodunos said they were ending negotiations with Watertight, who they said had failed to provide evidence of sufficient financial backing. That drew an angry
4	CBG	W_newsp_other_sports	also reported to have recruited a Norweigan dietary expert at a cost of 6,000 to provide evidence of his innocence but the expert is alleged to have disappeare
5	K32	W_newsp_other_sports	have to show they are going to the US only for the soccer matches and provide evidence of a steady history of employment, and that they can support themse
6	APN	W_ac_polit_law_edu	or not, environmental degradation is certainly happening. Thus we do not intend to provide evidence here of frightening rates of soil loss and desertification th
7	APN	W_ac_polit_law_edu	conservation policies. Thus, a rapid review of conservation world-wide can be expected to provide evidence as to whether there exist programmes which have
8	CCR	W_ac_polit_law_edu	more uniform patterns of working-class political activity (p 187). These do not provide evidence, however, of the efficacy of the national, rather of the converge
9	EDL	W_ac_polit_law_edu	penal or revenue law precluded the English court from giving the assistance requested; to provide evidence which is to be used in foreign proceedings, even e
10	FAM	W_ac_polit_law_edu	these examples of lack of public confidence in professional skill, expertise and judgement, provide evidence of what Habermas (1974,1975) has called the legit
11	FAM	W_ac_polit_law_edu	, and the Green Paper, The Development of Higher Education into the 1990s, provide evidence of a similar trend in higher education. The latter advocated the
12	FC7	W_ac_polit_law_edu	to the Hague Convention was in a position to give effect to it and to provide evidence to foreign courts in accordance with it without recourse to the court; and
13	FCK	W_ac_polit_law_edu	but only if and so far as compliance with the order of Buckley J. would provide evidence against him in a criminal trial. There is no reason why the privilege
14	FRB	W_ac_polit_law_edu	the party to press for more extreme policies. The years 1931,1951 and 1970 each provide evidence of such a movement; notably in the early fifties the Bevanit

Figure 14. Search results for provide in an academic discourse

12 Pair students and allow them to analyze concordance lines and understand their use in a differing context.
13 Ask students to click on ‘Collocates’ section in the middle of the page on the right.
14 Have students first type *provide* in the BNC, second choose ‘noun.ALL’ from the box on the second line, choose one word to the right by coloring ‘1’ to green on the right side, and last click on ‘Sections’ and choose ‘Spoken’ for the first box and ‘Academic’

for the second. Explain them that this would present the most frequent collocates with *provide* in the BNC in two discourses (see Figure 15).

Figure 15. Collocate search for the nouns immediately following provide in spoken and academic contexts

15 Ask students to hit the 'Find collocates' button in order to compare and contrast their occurrences in different discourses.

16 Ask students to analyze collocation frequency in spoken and academic discourses (see Figure 16) and give some implications. To exemplify;

   a *provide* occurs less in spoken discourse when it is compared to academic.

   b *provide* collocates with *information* in spoken discourse; on the other hand, in academic discourse it collocates with *evidence*.

British National Corpus (BNC)

SEARCH | FREQUENCY | CONTEXT | OVERVIEW

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (SPECIFIED SECTION) [HELP...]

SEC 1 (SPOKEN): 9,963,663 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1	INFORMATION	10	61	1.0	4.0	0.3
2	SERVICES	8	42	0.8	2.7	0.3
3	TRAINING	6	7	0.6	0.5	1.3
4	EMPLOYMENT	5	3	0.5	0.2	2.6
5	QUALITY	4	0	0.4	0.0	40.1
6	SITES	3	0	0.3	0.0	30.1
7	MONEY	3	2	0.3	0.1	2.3
8	FOOD	3	7	0.3	0.5	0.7
9	ADVICE	3	16	0.3	1.0	0.3
10	VALUE	2	0	0.2	0.0	20.1
11	RELIEF	2	3	0.2	0.2	1.0
12	RESOURCES	2	4	0.2	0.3	0.8
13	DETAILS	2	7	0.2	0.5	0.4
14	FUNDS	2	7	0.2	0.5	0.4
15	ASSISTANCE	2	12	0.2	0.8	0.3
16	MATERIAL	2	12	0.2	0.8	0.3
17	FACILITIES	2	13	0.2	0.8	0.2
18	SUPPORT	2	22	0.2	1.4	0.1
19	ASSISTANT	1	0	0.1	0.0	10.0
20	BONUS	1	0	0.1	0.0	10.0

SEC 2 (ACADEMIC): 15,331,668 WORDS

	WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	EVIDENCE	63	1	4.1	0.1	40.9
2	INFORMATION	61	10	4.0	1.0	4.0
3	SERVICES	42	8	2.7	0.8	3.4
4	SUPPORT	22	2	1.4	0.2	7.1
5	OPPORTUNITIES	18	0	1.2	0.0	117.4
6	CARE	17	1	1.1	0.1	11.0
7	SECURITY	16	1	1.0	0.1	10.4
8	ADVICE	16	3	1.0	0.3	3.5
9	DATA	14	0	0.9	0.0	91.3
10	FACILITIES	13	2	0.8	0.2	4.2
11	ANSWERS	12	0	0.8	0.0	78.3
12	ACCOMMODATION	12	1	0.8	0.1	7.8
13	ASSISTANCE	12	2	0.8	0.2	3.9
14	MATERIAL	12	2	0.8	0.2	3.9
15	ACCESS	11	0	0.7	0.0	71.7
16	REASONS	11	0	0.7	0.0	71.7
17	REMEDIES	10	0	0.7	0.0	65.2
18	EXAMPLES	9	0	0.6	0.0	58.7
19	EDUCATION	8	1	0.5	0.1	5.2
20	FEEDBACK	7	0	0.5	0.0	45.7

Figure 16. Words following provide in spoken and academic contexts (one word to the right)

17 Ask students to perform another collocate search for *give*.

18 Allow them to compare and contrast the collocates with *give* and give further implications (see Figure 17). To exemplify;

   a *give* collocates more with *evidence* in spoken discourse when it is compared to provide.

   b *give* collocates more with *rise* and *effect* in academic discourse.

British National Corpus (BNC)

SEARCH | FREQUENCY | CONTEXT | OVERVIEW

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (SPECIFIED SECTION) [HELP...]

SEC 1 (SPOKEN): 9,963,663 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1	WAY	53	45	5.3	2.9	1.8
2	EVIDENCE	32	88	3.2	5.7	0.6
3	PEOPLE	22	7	2.2	0.5	4.8
4	RISE	16	363	1.6	23.7	0.1
5	ADVICE	14	39	1.4	2.5	0.6
6	EFFECT	10	109	1.0	7.1	0.1
7	MUMMY	8	0	0.8	0.0	80.3
8	THANKS	7	1	0.7	0.1	10.8
9	CREDIT	7	9	0.7	0.6	1.2
10	MUM	6	0	0.6	0.0	60.2
11	TALKS	5	1	0.5	0.1	7.7
12	CONSIDERATION	5	3	0.5	0.2	2.6
13	BIRTH	5	15	0.5	1.0	0.5
14	PRIORITY	5	23	0.5	1.5	0.3
15	INFORMATION	5	27	0.5	1.8	0.3
16	MONEY	4	5	0.4	0.3	1.2
17	GUIDANCE	4	10	0.4	0.7	0.6
18	CHILDREN	4	15	0.4	1.0	0.4
19	MRS	3	0	0.3	0.0	30.1
20	THINGS	3	0	0.3	0.0	30.1

SEC 2 (ACADEMIC): 15,331,668 WORDS

	WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	RISE	363	16	23.7	1.6	14.7
2	EFFECT	109	10	7.1	1.0	7.1
3	EVIDENCE	88	32	5.7	3.2	1.8
4	REASONS	61	1	4.0	0.1	39.6
5	WAY	45	53	2.9	5.3	0.6
6	NOTICE	41	3	2.7	0.3	8.9
7	ADVICE	39	14	2.5	1.4	1.8
8	DIRECTIONS	28	1	1.8	0.1	18.2
9	INFORMATION	27	5	1.8	0.5	3.5
10	CONSENT	25	0	1.6	0.0	163.1
11	PRIORITY	23	5	1.5	0.5	3.0
12	EXAMPLES	18	1	1.2	0.1	11.7
13	CHILDREN	15	4	1.0	0.4	2.4
14	BIRTH	15	5	1.0	0.5	1.9
15	INSTRUCTIONS	13	0	0.8	0.0	84.8
16	SUPPORT	13	1	0.8	0.1	8.4
17	ATTENTION	12	2	0.8	0.2	3.9
18	PUPILS	11	0	0.7	0.0	71.7
19	GUIDANCE	10	4	0.7	0.4	1.6
20	ASSISTANCE	9	2	0.6	0.2	2.9

Figure 17. Words following give in spoken and academic contexts (one word to the right)

19 Give students a worksheet on academic and non-academic vocabulary to practice their knowledge.
20 Ask students to complete the task including verbs and their noun collocates.

4. Conclusion

This chapter suggests that corpus tools are useful resources in enhancing learners' vocabulary repertoire and presents some practical corpus-based activities for various levels of proficiency and ages. Considering that they are more effective than textbooks or dictionaries (Paker & Ergül-Özcan, 2017), corpus data can be used as resources that provide descriptive insights relevant to how language is used in authentic contexts and as tools that enable learners and teachers to identify linguistic patterns for pedagogical purposes. Moreover, corpus-based vocabulary activities attract learners' attention and thus make them hold positive attitudes towards such activities. Through concordances learners have instant access to the most requested learning materials. Once they learn how to use corpora, they will become more confident and autonomous in their learning progress.

REFERENCES

Ahsanuddin, M., Hanafi, Y., Basthomi, Y., Taufiqurrahman, F., Bukhori, H. A., Samodra, J., & Wijayati, P. H. (2022). Building a corpus-based academic vocabulary list of four languages. *Page Journal of Education and Instruction*, *12*(1), 159-167. https://doi.org/10.47750/pegegog.12.01.15

Alderson, J.C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. A&C Black. https://eclass.uoa.gr/modules/document/file.php/ENL286/Testing%20books/Diagnosing%20Foreign%20Language%20Proficiency_Alderson.pdf

Aston, G. (2001) Learning with corpora: An overview. In G. Aston (Ed.), *Learning with corpora* (pp. 7–45). Athelstan.

Bernardini, S. (2002). Exploring new directions for discovery learning. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis: Proceedings of the Fourth International Conference on Teaching and Language Corpora* (pp. 165-182). Rodopi.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, *67*(2), 348–393. https://hal.archives-ouvertes.fr/hal-01850690/document

Cobb, T. (1999). Applying constructivism: A test for the learner-as-scientist. *Educational Technology Research and Development*, *47*(3), 15–31. https://doi.org/10.1007/BF02299631

Cobb, T., & Boulton, A. (2015). Classroom applications of corpus analysis. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 478-497). Cambridge University Press. https://doi.org/10.1017/CBO9781139764377.027

Cortes, V. (2007). Exploring genre and corpora in the English for academic writing class. *The ORTESOL Journal*, *25*(8), 14. https://www.researchgate.net/publication/273379093_Exploring_genre_and_corpora_in_the_English_for_academic_writing_class

Cotos, E. (2014). Enhancing writing pedagogy with learner corpus data. *ReCALL*, *26*(2), 202-224. https://doi.org/10.1017/S0958344014000019

Daskalovska, N. (2015) Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, *28*(2), 130-144. https://doi.org/10.1080/09588221.2013.803982

The Corpus of Contemporary American English: 450 million words. http://corpus.byu.edu/coca/

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, *14*, 159–90. https://doi.org/10.1075/ijcl.14.2.02dav

Frankenberg-Garcia, A. (2014). The use of corpus examples for language comprehension and production. *ReCALL*, *26*(2), 128–146. https://doi.org/10.1017/S0958344014000093

Gilquin, G., Granger, S., & Paquot, M. (2007) Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, *6*(4), 285–374. https://doi.org/10.1016/j.jeap.2007.09.007

Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 359-370). Routledge. https://doi.org/10.4324/9780203856949

Hart, B., & Risley, T.R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.

Hinkel, E. (2011). What research on second language writing tells us and what it doesn't. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 523-538). Routledge.

Hu, M., & Nation, I.S.P. (2000). Unknown vocabulary density and reading comprehension. *Reading in A Foreign Language*, *23*, 403–430. https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/c5873d5c-23b5-41d1-99a5-fde539883ceb/content

Johns, T. (1988). Whence and whither classroom concordancing? In T. Bongaerts, P. de Haan, S. Lobbe & H. Wekker (Eds.), *Computer applications in language learning* (pp. 9–27). Foris. https://doi.org/10.1515/9783110884876

Kiliç, M. (2019). Vocabulary knowledge as a predictor of performance in writing and speaking: A case of Turkish EFL learners. *PASAA: Journal of Language Teaching and Learning in Thailand*, *57*, 133-164. https://files.eric.ed.gov/fulltext/EJ1224421.pdf

Kucera, H., & Francis, W.M. (1967). *Computational analysis of Present-Day American English*. Brown University Press. https://www.jstor.org/stable/1263890

Lee, C., Ge, H., & Chung, E. (2021). What linguistic features distinguish and predict L2 writing quality? A study of examination scripts written by adolescent Chinese learners of English in Hong Kong. *System*, *97*, 102461. https://doi.org/10.1016/j. system.2021.102461

Liu, D., & Lei, L. (2018). Using corpora to teach vocabulary. In J. Liontas (Ed.), *TESOL Encyclopedia of English Language Teachi*ng (pp. 3438–3445). Wiley-Blackwell.

Ma, Q., Tang, J., & Lin, S. (2021). The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online

collaboration. *Computer Assisted Language Learning*, 1–30. https://doi.org/10.1080/09588221.2021.1895225

McEnery, T., & Wilson, A. (1997). Teaching and language corpora. *ReCALL*, 9(1), 5–14. https://doi.org/10.1017/S0958344000004572

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Routledge.

Moon, R. (2010). What can a corpus tell us about lexis? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 197-211). Routledge. https://doi.org/10.4324/9780203856949

Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art and beyond. In S. Braun, K. Kohn & J. Mukherjee (Eds.), *Corpus technology and language pedagogy* (pp 5-24). Peter Lang. https://www.uni-giessen.de/faculties/f05/engl/ling/staff/professors/jmukherjee/publications/pdfs/Mukherjee2006_pedagogy.pdf

Mueller, C.M., & Jacobsen, N.D. (2015). A comparison of the effectiveness of EFL students' use of dictionaries and an online corpus for the enhancement of revision skills. *ReCALL*, *28*(1), 3-21. https://doi.org/10.1017/S0958344015000142

Nation, I.S.P., & A. Coxhead (2021). Vocabulary learning and teaching. In P. Szudarski & S. Barclay (Eds.), *Vocabulary theory, patterning and teaching* (pp. 197–219). Multilingual Matters.

Nouri, N., & Zerhouni, B. (2016). The relationship between vocabulary knowledge and reading comprehension among Moroccan EFL learners. *Journal of Humanities and Social Science*, *21*(10), 19-26. https://dx.doi.org/10.9790/0837-2110051926

O'Keeffe, A., & Farr, F. (2003). Using language corpora in initial teacher education: Pedagogic issues and practical applications. *TESOL Quarterly*, *37*(3), 389–418. https://doi.org/10.2307/3588397

Oxford University Computing Services. (1995). *The British National Corpus*. Oxford University Press.

Paker, T., & Ergül-Özcan, Y. (2017). The effectiveness of using corpus-based materials in vocabulary teaching. *International Journal of Language Academy*, *5*(1), 62-81. http://dx.doi.org/10.18033/ijla.3494

Pérez-Paredes, P., & Cantos Gómez, P. (2004) Some lessons students learn: Self-discovery and corpora. In G. Aston, S. Bernardini & D. Stewart (Eds.), *Corpora and language learners* (pp. 247–257). John Benjamins. https://doi.org/10.1075/scl.17.18per

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press. http://assets.cambridge.org/052166/0483/sample/0521660483wsn01.pdf

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, *95*(1), 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Seidlhofer, B. (2002). Pedagogy and local learner corpora: Working with learning-driven data. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 213–234). Benjamins. http://ndl.ethernet.edu.et/bitstream/123456789/57648/1/183pdf.pdf#page=224

Sripicharn, P. (2010). How can we prepare learners for using language corpora? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 371-384). Routledge. https://doi.org/10.4324/9780203856949

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honor of H. G. Widdowson* (pp. 125–144). Oxford University Press.

Szudarski, P. (2018). *Corpus linguistics for vocabulary: A guide for research*. Routledge.

Thurstun, J., & Candlin, C.N. (1998). Concordance and the teaching of the vocabulary of academic English. *English for Specific Purposes*, *17*, 267–280. https://doi.org/10.1016/S0889-4906(97)00013-6

Uchihara, T., & Saito, K. (2019). Exploring the relationship between productive vocabulary knowledge and second language oral ability. *Language Learning Journal*, *47*(1), 64-75. https://doi.org/10.1080/09571736.2016.1191527

Wang, Y., & Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: The contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System*, *65*, 139–150. https://doi.org/10.1016/j.system.2016.12.013

Wilkins, D. A. (1972). *Linguistics in language teaching* (Vol. 111). Edward Arnold.

Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology*, *12*(2), 31–48. https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/9d7d4802-8866-4246-af19-69187c5f14bc/content

Yoon, H., & Hirvela, A. (2004). E.S.L. student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, *13*(4), 257–283. https://doi.org/10.1016/j.jslw.2004.06.002

Zhang, S., & Zhang, X. (2022). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research*, *26*(4), 696-725. https://doi.org/10.1177/1362168820913998

To Cite this Chapter:

Erdemir, N. & Yeşilçınar S. (2022). How to use corpora in language teaching for vocabulary enhancement. In A. Önal & K. Büyükkarcı (Eds.), *Essentials of foreign language teacher education*, (pp. 64-85). ISRES Publishing.

ABOUT THE AUTHORS

Dr. Nihan ERDEMİR

ORCID ID: 0000-0002-8610-3590

nihanerdemir@gmail.com

Suleyman Demirel University

Nihan Erdemir is an English Language Teaching Assistant Professor at Süleyman Demirel University, Turkey. She holds a Ph.D. from Gazi University, Turkey, and an M.A. from the University of Vienna, Austria. She taught English and Turkish in Austria and Turkey for over seven years. Her research interests are academic writing, language teacher education, and classroom research.

Dr. Sabahattin YEŞİLÇINAR

ORCID ID: 0000-0001-6457-0211

s.ycinar@alparslan.edu.tr

Muş Alparslan University

Sabahattin Yeşilçınar is an Assistant Professor of English Language Teaching at Muş Alparslan University, Turkey. He holds a Ph.D. from Gazi University, Turkey, and an M.A. from Yüzüncü Yıl University, Turkey. His research primarily deals with teacher education, teacher assessment, and technology in language teaching and learning