https://doi.org/10.17323/jle.2023.12426

The Effectiveness Data-Driven Vocabulary Learning: Hands-on Concordancing through a Pedagogical Corpus

Sibel Tosun [1], **Hatice Sofu** [2]

[1] Firat University, Elazig, Turkey

[2] Cukurova University, Adana, Turkey

ABSTRACT

Introduction: Although extensive studies have been carried out on the effectiveness of corpora on teaching vocabulary, the exploration of whether learners can benefit from a pedagogical corpus, particularly regarding hands-on engagement by lower-level learners, has received little attention.

Purpose: To address this gap in the literature, this study sets out to explore the effectiveness of Data-Driven Learning (DDL) in enhancing the vocabulary acquisition of EFL students at a state university in Turkey through a pedagogical corpus.

Method: The quasi-experimental study employed a mixed-method research design, in which both quantitative and qualitative data were gathered through vocabulary tests, student questionnaires, and semi-structured interviews. Fifty-eight low-level students with an average age of 19 served as participants. The experimental group made use of hands-on concordancing while the control group received conventional course book-based instruction to learn the target words.

Results: The results indicate that pedagogical corpora have significant potential in facilitating vocabulary learning of low-level learners. The vocabulary tests revealed that the students who practiced with DDL outperformed the students who received traditional vocabulary instruction in both the post-test and the delayed post-test. The findings from student questionnaires, and semi-structured interviews also denoted that the participants held positive attitudes towards using concordancing to expand their vocabulary and grow aware of some aspects of words such as part of speech information, different meanings and usages, lexico-grammatical structures, and collocations.

Conclusion: The present study provides useful implications for collection and use of a pedagogical corpus for classroom use.

KEYWORDS

corpus-based instruction, data-driven learning (DDL), concordance lines, vocabulary learning, pedagogical corpus

Citation: Tosun S., & Sofu H. (2023). The Effectiveness Data-Driven Vocabulary Learning: Hands-on Concordancing through a Pedagogical Corpus. *Journal of Language and Education, 9*(3), 176-190. https://doi.org/10.17323/jle.2023.12426

Correspondence:
Sibel Tosun,
sibeltosun@outlook.com

Received: May 18, 2021
Accepted: September 15, 2023
Published: September 30, 2023

CC BY

INTRODUCTION

Considering the impact corpora have had on language pedagogy, they are acknowledged to have "revolutionized" language teaching in various areas (Conrad, 2000, p. 549) both directly and indirectly (Römer, 2008). Dictionary making (e.g., Gouws, 2021; Hunston, 2002; O'Keefe et al., 2007), textbook and material development (e.g., Boulton, 2012; Friginal & Roberts, 2022; McEnery & Xiao, 2011), design of syllabi and testing materials (e.g., Boulton 2009; Hunston, 2002) including their validation and standardization (McEnery & Xiao, 2011) are some areas where corpora have indirectly contributed to language teaching. A resurgence of interest has also been witnessed in incorporating corpora into language teaching directly (Römer, 2008) through DDL, an approach to language learning introduced by Johns (1991).

A myriad of studies focused on the impact of the DDL and corpus use in L2 learning. In particular, recent meta-analyses (Boulton & Cobb 2017; Cobb & Boulton, 2015; Lee et al., 2019), have put forward the positive outcomes reported by corpus studies. With regard to teaching language skills, some investigated the role of corpora and concordancing in writing (Gilmore, 2009; Huang, 2014), and some in grammar (Boulton, 2009; Girgin, 2011; Vannestal and Lindquist, 2007). Although a considerable number of studies exist in literature to date, the studies on the effectiveness of corpus use in vocabulary learning based on students' performance (Boulton, 2008, 2012; Chan & Liou, 2005; El-Esery, 2015; Gilmore, 2009; Golabi, 2022; Kazaz, 2015; Koosha & Jafarpour, 2006; Lee & Lin, 2019) have focused on the learning of intermediate/advanced level students. Since existing corpora mostly appeal to high-level learners (Flowerdew, 2012, Meunier, 2011), the research on lower-level students' vocabulary learning remained relatively few. It is possible that researchers have tended to work with higher level students because native corpora have a high vocabulary load (Balunda, 2009, Sinha, 2021), and are not considered learner-centered, and culturally and contextually appropriate for all (Meunier, 2011). This could be one of the reasons as to why corpora use has not been normalized in language teaching and learning, and research-practice gap still exists (Chambers, 2019). To bridge this gap, the creation of a contextually relevant pedagogical corpus (as suggested by Braun, 2005) emerges as a potential solution, particularly in addressing challenges associated with the application of general corpora, especially in terms of learner-corpus interaction complexities (Kavanagh, 2021). Especially promising for learners at lower proficiency levels, who primarily engage with classroom language and struggle to comprehend less common vocabulary within native corpora, a pedagogical corpus offers the potential to amplify benefits.

Central to this study is the exploration of whether learners can benefit from a pedagogical corpus, a research that has received limited attention, particularly regarding hands-on engagement by lower-level learners with such a resource. Therefore, the novelty of this research lies in its attempt to employ a pedagogical corpus that comprises textbooks and teaching materials for vocabulary instruction. With this objective, the study seeks to investigate the efficacy of Data-Driven Learning (DDL) in enhancing students' vocabulary acquisition through the utilization of a pedagogical corpus.

LITERATURE REVIEW

Data-Driven Learning

Tim Johns's concept of DDL (1991) is a method of language learning that involves the exploration of regularities and patterns in language samples. Johns's earlier attempts to incorporate corpora into language teaching, as well as his seminal publications (1986, 1988, and 1991), and his website has become the backbone of subsequent pedagogical applications of corpora and DDL. One reason for the attention DDL received at the time was that it brought a new perspective to language teaching and learning. For example, it differs from traditional teaching methods in terms of the nature of instruction and the roles of students and teachers. It provides learners with the opportunity to examine corpora and recognize language patterns (Boulton, 2012; Hunston, 2002; O'Keeffe et al., 2007; Friginal & Roberts, 2022) enhances their ability to make generalizations (Gilquin, 2021; Johns, 1991), and thereby fostering inductive learning (O'Keeffe et al., 2007). In this regard, it also has the potential to enhance autonomous learning (Barabadi & Khajavi, 2017; Binkai, 2012).

As a prominent "computing tool for the data-driven approach" (Johns, 1991, p. 2), concordancers can analyze and organize massive amounts of texts in a very short time and present "potential patterns" by reducing all language data to a simple list in alphabetical order (Scott & Tribble, 2006, p. 5). Concordance lines and the keyword-in-context (KWIC) format facilitate the process of identifying patterns such as "lexical features" and "phraseology" of words (Sripicharn, 2003, p. 204) not only quantitatively but "beyond the frequency of the words" (O'Keeffe et al., 2007, p. 2).

Vocabulary Teaching Through DDL

While teaching English in schools, irrespective of whether it is presented under integrated skills courses or separately, traditional methods are commonly used to teach vocabulary such as providing synonyms, antonyms, L1 translations and definitions (Balcı & Çakır, 2011). Traditional vocabulary learning is usually characterized as an inactive process for learners as they usually obtain word lists and try to memorize them (Chen, 2004). However, memorizing does not necessarily mean that acquisition takes place (Nation, 2001). Understanding the form, meaning and usage of words (Nation, 2001) and knowledge of surrounding vocabulary are also crucial for acquisition (Harmer, 1993). In this respect, DDL is a potent alternative to traditional methods as it guides learners to observe and analyse data, look for patterns implicitly, and become autonomous researchers.

DDL is considered a viable approach to vocabulary learning on various grounds. First, it serves as an opportunity to get exposed to words in different contexts and forms (Barabadi & Khajavi, 2017; Wu et al, 2010), which in turn facilitates vocabulary expansion (Nation, 2001). Moreover, it increases both vocabulary breadth and depth as vocabulary items are presented in their collocative environment (Chen, 2004), and in their most common forms and patterns, which provides a means of studying collocates (McEnery & Xiao, 2011; Varley, 2009). On top of that, the above-specified vocabulary gains are not temporary as research (El-Esery, 2015, Karras, 2015) suggests that DDL yields long-term benefits and better vocabulary retention compared to traditional vocabulary

learning. DDL research (Binkai, 2012; Varley, 2009) revolved around immediate student attitudes and perceptions, while fewer studies focused on learners' performance (Barabadi & Khajavi, 2017; El-Esery, 2015, Kazaz, 2015). Furthermore, conducting empirical research on DDL offers limited results due to their tendency to focus on "specific, immediate learning outcomes", which makes it difficult to examine long-term achievement (Boulton, 2010, p.536). Therefore, conducting more empirical research on the long-term effects of DDL seems to be focal for our increased understanding of vocabulary retention through DDL practices.

Although its merits have been amply acknowledged in the existing literature, corpus resources and tools have not gained widespread adoption among language teachers and learners in the language classroom (Pérez-Paredes, 2022). Corpus use in language pedagogy is not without its criticism. Conversing views on the suitability of corpora for low-level students exist. While acknowledging that corpora such as the BNC appeal to advanced-level users, Aston[1] asserts that learners do not need to make sense of all data in corpus samples. On the issue, Flowerdew (2012) maintains that mostly higher-level learners can make sense of corpora while Lee & Liou (2003) found that low-level learners can benefit from corpora more compared to high-level learners. Conceding the difficulty of corpus samples, Chen (2004) suggests preparing concordance sheets to avoid dealing with irrelevant data. Toriida (2016) further recommends that teachers compile their corpus from textbooks, readers and journal articles for learners.

Despite being authentic in nature, the authenticity of native corpora has been questioned for learners from another culture, as well as their effectiveness in language learning (see, e.g., Widdowson, 2003). Drawing attention to the pedagogically relevant corpora, Braun (2005) postulates that authenticity can be improved to some extend if a corpus is pedagogically relevant in terms of content, language and culture. Textbook corpora and pedagogical corpora emerge as two important candidates at this point. As an extended version of textbook corpora, Meunier and Gouverneur (2009) define pedagogic corpora as "large enough and representative sample of the language, spoken and written, a learner has been or is likely to be exposed to via teaching material, either in the classroom or during self-study activities" (p.186). The terms *pedagogic corpora* and *pedagogical corpora* were used later in works of other scholars (see, e.g., Bennet, 2010; Chambers 2019), and such corpora have previously been created and exploited for textbook analysis (Bergström et al., 2023; Chen & Yuhua, 2023; Meunier, & Gouverneur, 2009; Sun & Dang, 2020), and writing genre analysis (Melissourgou & Frantzi, 2019). However, raising awareness of learners of all the occurrences of a word in different contexts (Huston, 2002) is an underrecognised exploitation of pedagogic corpora. Thus, taking the aforementioned issues into account, this study explored the effectiveness of DDL on vocabulary learning of tertiary-level EFL students by utilizing a pedagogical corpus. Two research questions guided the study:

(1) Is there a statistically significant improvement in vocabulary knowledge of target words for the participants who practiced through DDL when compared to those who received traditional vocabulary instruction through textbook materials?

(2) What are the attitudes of the DDL group towards learning vocabulary through DDL?

METHOD

Setting & Participants

The present study was undertaken at the School of Foreign Languages, at a state university in Turkey. At this institution, learners who cannot meet the language requirements to study in their respective departments receive general English language skills instruction. Through a one-year academic program, language skills are presented in an integrated way through "Main Course" and other supporting skill courses such as Listening & Speaking, and Reading & Writing. However, there is no special course devoted to vocabulary instruction.

Fifty-eight Turkish students (M = 44, F = 14) whose ages ranged between 18 and 22 took part in the study. The participants were placed in elementary classes based on the results of the placement test administrated at the beginning of the year and were receiving pre-intermediate level instruction during the study.

In this quasi-experimental study with a comparison group pre-test/post-test design, the participants were divided into experimental (N = 29) and control groups (N = 29) based on convenience sampling; that is, four classes with almost the same number of students were selected. Two intact classes were assigned to the experimental group, and two other intact classes formed the control group. To control some confounding variables such as course delivery, course content and duration, two instructors made sure that both groups followed exactly the same curriculum, courses and materials as planned by the administration and always updated each other for specific classroom practices.

Corpus Compilation

Before the intervention, a pedagogical corpus was compiled by the researchers with reading texts learners can

[1] Aston, G. (1998). *Learning English with the British National Corpus*. [Paper presentation]. 6th Jornada de Corpus, UPF, Barcelona. https://www.sslmit.unibo.it/~guy/barc.htm

make sense of while doing their analysis. For this reason, advanced-level texts were excluded. Samples of reading texts from elementary, pre-intermediate and intermediate level course-books and listening scripts from websites (see appendix A for the list of the sources) were included in the data set to compile a sample of both written and spoken language. The corpus contained a total of 358,972 tokens and 14,402 word types. While 27% of the corpus consisted of spoken language scripts, 63% consisted of written texts.

Classroom Materials

For the participants in the experimental group, researchers created five paper-based concordance handouts, as lower-level learners require guidance and support from their teachers to navigate the data they encounter during vocabulary searches (Granath, 2009). The handouts were developed following Nation's (2001) categorization of aspects of word knowledge: form, meaning and use. The questions in the handouts were constructed to raise participants' awareness of word forms, prefixes and suffixes, form-meaning relationships, collocations, and word usage (see appendix B for a sample concordance handout).

The aim was to introduce five target vocabulary items in each session. The target words were selected from the scheduled units in the Reading & Writing coursebook, with which the participants were being instructed. An adapted version of the *Vocabulary Knowledge Scale* (Schmitt & Zimmerman, 2002) was administrated to select less familiar vocabulary items among 36 pre-selected target words. The participants rated their own vocabulary knowledge with a four-level scale. 25 words which held the highest-rated "I do not know the word" option were selected as final target words.

For the exploration of the pedagogical corpus, the participants used the computer software "Antconc 3.4.4". The researchers downloaded the software for students' use before the intervention.

Data Collection Tools

Quantitative data was collected through a vocabulary test, which was administrated at three time points as pre-test, post-test and delayed post-test. Since an experimental design with a control group is adopted for the study, the existence of a control group reduces the test effect, in that the differences in the groups are not attributed to the retesting effect, but more likely to the intervention (Cook & Campbell, 1979). The vocabulary test, which was worth 100 points, comprised four parts: a) seven multiple-choice questions b) seven paragraph gap-filling questions c) six questions that require filling in concordance lines with target words d) five sentence-matching questions. Two academics holding M.A. in ELT examined the test in terms of content and face validity. They particularly checked whether the test includes all target words, and whether it measures different aspects of word knowledge specified in the objectives. They also evaluated the test with relation to level-appropriateness, clarity and formatting. Later, a pilot study was conducted with 20 students who were on the same level as the target population. The Cronbach Alpha was calculated as .94, which indicated high internal consistency.

A small questionnaire was also developed to explore students' perceptions of DDL after the intervention. It comprised 10 items in the format of a 5-point Likert Scale. While items 1 and 2 were created by the researchers, the remaining items were adapted from Boulton (2010), Girgin (2011), and Jablonkai & Čebron (2017). The reliability of the questionnaire was checked after data collection since it was only targeted at the research sample, and it was found to have internal consistency, r = .720.

Semi-structured focus group interviews were also conducted with seven volunteers from the experimental group to further investigate the efficacy of data-driven vocabulary learning, and explore opinions and experiences. Interviews were conducted after the intervention, and in the participants' mother tongue.

The Data Collection Procedure and Data Analysis

The study was conducted within three months in Spring Semester, 2016. One week prior to the pre-test, the Vocabulary Knowledge Scale was administered, and the final decision was made on the target words to be introduced. For the pre-test, which was given to both groups on the same day, the participants were given 30 minutes. Following that, the experimental group was given an introductory session on how to use the concordance program, open corpus files, and search for the target words.

During five-week treatment, the participants in the experimental group received 12 sessions that took 70-80 minutes each. In each session, they received a concordance handout, and were asked to analyse the corpus data following the "Identify - Classify - Generalise" procedure specified by Johns (1991, p. 5). Within this period, which took about 45-50 minutes of the session, the students were asked to identify relevant concordance lines, classify the patterns, and make generalizations to answer the questions in their handout. In the meantime, the instructing researcher observed the class and answered student questions. In the last 20-25 minutes of the session, a whole class discussion was held to elicit correct responses, alternative sentences or phrases identified from the concordance lines, and to uncover what they had learned. On the other hand, the control group covered their coursebook according to the schedule, and were introduced to the target vocabulary items through explicit teaching of teacher explanations and coursebook activities.

Upon completion of the treatment, the immediate post-test was administered to both groups in the same week. The experimental group was given the attitude questionnaire immediately after the post-test. The following week, seven students from the experimental group were interviewed. Five weeks after the immediate post-test, the participants were also given the delayed post-test.

For data analysis, the quantitative data collected through the vocabulary tests and the questionnaire were analysed using the Statistical Package for the Social Science (SPSS). On the other hand, the qualitative data, which were collected through interviews, were analysed using Creswell's (2012) qualitative content analysis scheme. The data were first transcribed, and the codes were created based on the recurring statements. Later, the themes that emerged from the codes were labelled.

RESULTS

The Effectiveness of DDL Based on the Vocabulary Tests

In an attempt to answer the first research question, which aims to find out if the participants show any improvement in their vocabulary knowledge of the target words, a descriptive analysis was first computed for the pre-test, post-test, and delayed post-test of the control and the experimental group (DDL group).

Table 1 illustrates that the control group (M = 40.76, SD = 21.1) scored 3 points higher than the experimental group (M = 37.8, SD = 18) in the pre-test. The mean scores for the immediate post-test indicate that although both groups showed improvements, the participants in the DDL group (M = 68.96, SD = 19.02) outperformed the control group (M = 53.85, SD = 20.29). Regarding the scores of delayed post-test, which was administrated five weeks after the post-test, the scores of both groups decreased compared to their post-test scores. However, the mean score for the control group (M = 48.23, SD = 22.47) dropped slightly more than the experimental group (M = 64.67, SD = 23.05).

Before comparing results for any significance, the baseline scores of participants in each group were first compared. The data was normally distributed and homogenous, so independent samples t-test was performed. The results presented in Table 2 indicate that although mean pre-test score of the control group was 2.93 points higher than experimental group's mean, these means did not differ significantly between the experimental group (M = 37.8, SD = 18) and the control group (M = 40.76, SD = 21.1) ([t [51] = -.54, p = 0.591]). Therefore, it can be concluded that the groups were not significantly different prior to intervention.

In order to explore the impact of concordance training on the vocabulary knowledge of target vocabulary items in three time intervals, 2x3 Mixed ANOVA was also performed. First, the assumptions of homogeneity of variances were tested and satisfied based on Levene's F test for the pre-test (F(1, 51) = 1.32, p = .256), post-test scores (F(1, 51) = 1.12, p = .30) and delayed post-test scores (F(1, 51) = 0.65, p = .80). Mauchly's test indicated that the assumption of sphericity has not been met, χ 2 (2) = 7.94, p = .020. Hence, degrees of freedom were corrected using Huynh-Feldt estimates of sphericity (ε = .91).

Table 1

Descriptive Statistics for the Test scores of Control and Experimental Groups

Tests	Groups	*N*	*M*	*SD*
Pre-test	Experimental	27	37.85	18.06
	Control	26	40.77	21.13
Post-test	Experimental	27	68.96	19.02
	Control	26	53.85	20.29
Delayed post-test	Experimental	27	64.67	23.05
	Control	26	48.23	22.47

Table 2

Independent Samples T-Test for Pre-Test Scores of the Control and Experimental Group

	Levene's Test for Equality of Variances		T-test for Equality of Means				
	F	Sig.	t	df	p.	Mean Dif.	Std. Error Dif.
Pre-test	*1.32*	*.256*	*-.54*	*51*	*.591*	*-2.93*	*5.39*

Table 3

Within Subject Effects for the Control and the DDL group

Source		Type III Sum of Squares	*df*	Mean Square	*F*	Sig.	Partial Eta Squared
time	Sphericity Assumed	14654.92	2	7327.46	51.16	.000	.51
	Huynh-Feldt	14654.92	1.83	7994.22	51.16	.000	.51
time*group	Sphericity Assumed	3390.92	2	1695.46	11.84	.000	.19
	Huynh-Feldt	3390.92	1.83	1849.74	11.84	.000	.19

Table 4

Pairwise Comparisons of Test Scores across the Control and the DDL Group

Group	(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
experimental	1	2	-31.11	3.57	.000	-38.28	-23.94
		3	-26.81	3.63	.000	-34.10	-19.53
	2	3	4.30	2.53	.096	-.792	9.38
control	1	2	-13.08	3.64	.001	-20.38	-5.76
		3	-7.46	3.70	.049	-14.89	-.034
	2	3	5.61	2.58	.034	.430	10.80

As shown in Table 3, the main effect of time (pre-test, post-test and delayed post-test) on vocabulary test scores was significant, Huynh-Feldt, $F(1,91.66) = 51,16$, $p = <.001$, with a large effect (Eta-squared = .51), indicating that there were changes over time in vocabulary scores across the whole sample. This effect was qualified with a significant time and group (experimental/control) interaction effect, Huynh-Feldt, $F(1.83, 91.66) = 11.84$, $p = <.001$, eta squared = .19 revealing that the changes in vocabulary scores of participants over time are not equivalent across the two groups. This indicates that the groups are changing in different ways.

Post-hoc pairwise comparisons were also performed using paired-samples t-test to determine the levels of significance.

Follow-up pairwise comparisons presented in Table 4 indicated that the vocabulary scores of the treatment group significantly increased from time 1 to time 2 ($p = <.001$, cohen's $d = 1.68$), and time 1 to time 3 ($p = <.001$, $d = 1.29$) with a large effect size. However, no statistically significant difference was found between time 2 and time 3 ($p = .096$, $d = .20$), indicating that although the vocabulary scores have decreased from time 2 to 3 in the experimental group, it was not statistically significant. As for the control group, there was a statistically significant difference between time 1 and 2 ($p = .001$, $d = .63$) with a medium effect size, and time 1 and time 3 ($p = .049$, $d = .38$) with a small effect size. On the other hand, there was a significant decrease in the vocabulary scores from time 2 to 3 ($p = .034$, $d = .21$), which showed a rather poor retention rate in the vocabulary scores. These findings indicate that the experimental group showed a significantly greater improvement in the post test compared to the control group. Although both groups obtained lower scores in the delayed post-test than the post-test, the decrease in the scores is significant only in the control group, which implies that there is an overall higher improvement in the DDL group.

Learner Attitudes towards DDL

The effectiveness of DDL was further explored with an attitude questionnaire and focus group interviews. The findings regarding the attitude questionnaire are presented in Table 5.

As indicated in Table 5, overall, learners seemed to have a positive attitude towards DDL and believed that it improved their English. The mean score for item 4 was the highest ($M = 4.29$, $SD = .66$), showing that a big portion of students thought that studying concordance lines had a facilitating role in increasing knowledge of word usage. The lowest mean score, on the other hand, belonged to item 6 ($M = 2.07$, $SD = .91$), which indicated learners' disagreement on the difficulty of concordance lines. The participants found

DDL practices enjoyable (*M* = 3.96, *SD* = .70), wanted to continue studying with concordance lines (*M* = 4.03, *SD* = .75), and suggested that instructors utilize concordance to teach vocabulary (*M* = 4.11, *SD* = .93).

The findings from the interviews align with the results of vocabulary tests and the attitude questionnaire. The emerging themes and codes were listed in Table 6. The frequency of the codes expressed by participants is presented in parentheses.

The interview analysis revealed three themes that reflect the participants' experiences with data-driven learning. Overall, all the participants mentioned that they had fun while studying through concordance lines emphasizing its dissimilarity to other vocabulary learning methods they know. Some stated that they enjoyed studying vocabulary since it was computer-based and had a colourful display. One student remarks:

> *I had fun learning vocabulary with concordance lines because in class, we generally write or read, but here we did something different with computers and it was not boring at all.*

Regarding the strengths of data-driven learning, which emerged as the second theme, all the participants noted that they had the chance to analyse words in different sentences, and understand how words were used in those sentences, as seen in the extracts below:

> *There were only one or two example sentences in textbooks, but here I have seen a lot of sentences. I have also learned other forms of words.*

> *With the help of concordance lines, I have learned how words are presented in different forms such as verbs, nouns adjectives, etc., and I have also learned which prepositions are used after certain words because they were easy to see with concordance lines.*

The statements on how concordance lines helped students learn more words than they intended are worth attention. In particular, participants mentioned gains in word forms, collocations and lexico-grammatical structures. Some also pointed out the permanency of the knowledge acquired through data-driven learning. The following quotations show in what ways participants benefited from data-driven learning:

> *Concordance lines helped me learn vocabulary because we learned how the words we come across in the classroom are used in sentences, and their position in those sentences... while doing an activity, you learn a lot, for example, the prepositions, and the words with different meanings.*

> *Words in coursebooks are restricted, but here we can learn more words with concordance lines.*

> *Colourful words helped me a lot to identify which words are nouns, adjectives, verbs, etc. I sometimes remember the words with their colours. It became easier for me to remember them.*

The last theme that was drawn from student interviews was the concerns related to DDL practice. Two participants stated that doing research seemed complicated at first, yet they immediately added that they figured it out in a short time. Four participants were also cautious about using only concordance lines for vocabulary learning and suggested learning through concordance lines along with coursebook instruction. One participant touched upon this with the following extract:

> *I think that learning through concordance lines is beneficial, but it would be more effective when used with activities in our coursebook. I think they support each other.*

Based on the findings, we may conclude that data-driven learning had a positive impact on the vocabulary learning of the learners especially in the areas such as vocabulary expansion, word usage, word formation and collocations.

Table 5

Mean Scores for All the Items of the Attitude Questionnaire

	Items	*M*	*SD*
I think....			
1.	studying vocabulary through concordance lines is enjoyable.	3.96	.70
2.	studying vocabulary through concordance lines helps improve my English.	4.25	.59
3.	using concordance lines improved my English writing ability	3.74	.71
4.	using concordance lines is helpful for learning the usage of vocabulary.	4.29	.66
5.	using concordance lines in the learning of English vocabulary increased my confidence in learning English vocabulary.	3.29	.72
6.	learning vocabulary through concordance lines is more difficult than learning vocabulary through a coursebook.	2.07	.91
7.	learning vocabulary through concordance lines is more boring than learning vocabulary through a coursebook.	2.88	1.21
8.	I prefer using concordance lines in learning of English vocabulary to using a coursebook in learning of English vocabulary.	3.51	1.12
9.	I recommend that teachers should use concordance lines so as to teach vocabulary in EFL classes.	4.11	.93
10.	I would like to do more concordance activities in class.	4.03	.75

Table 6

Students' Perceptions on the Effectiveness of Data-Driven Learning

Themes	Codes
Affective factors	Fun (x7)
	Attractive due to being different (x6)
	Colourful layout (x5)
	Enjoyable since computer-based (x4)
Strengths	Multiple sample sentences (x7)
	Word forms (x7)
	Word usage (x7)
	Collocations + lexico-grammatical structures (x6)
	Learning beyond intended (x4)
	Permanent (x4)
Concerns	Analysing seemed a bit complicated at first (x2)
	A combination of coursebook and concordance can be more effective (x4)

DISCUSSION

The present study sought to explore the efficiency of the data-driven approach on EFL students' vocabulary learning. The results indicated that although both groups demonstrated learning of the target words, the DDL group showed a greater improvement and performed significantly better than the coursebook group on the post-test. The observed improvements in both groups were no surprise since the target vocabulary items were unknown to the participants before the study, and after the instruction period, it was expected that both groups would show improvement to some extent. However, the gain was higher by the DDL group, which indicated that data-driven practice, enriched with corpus and concordance was more efficient in enhancing students' vocabulary knowledge. The results echo previous empirical studies (see, for example, Binkai, 2012; El- Esery, 2015; Koosha & Jafarpour, 2006; Karras, 2015 and Kazaz, 2015), which reported that corpus-based learning yielded better results than other methods of vocabulary instruction.

The results regarding the delayed post-test revealed that the DDL group had a higher retention rate although both groups showed a decrease in the test scores, which indicated that DDL was more beneficial for the learners than the conventional coursebook-based instruction in the long run as well. The reasoning behind better retention could be drawn from the interviews in that some learners linked remembering words for a longer time to various examples, concordance display and colourful layout, supporting Schmitt (2008), who highlighted that exposure coupled with attention directed to lexical items results in better attention.

As regards to the learners' perceptions, the analysis of the questionnaires and interviews reflected positive attitudes towards DDL. This supports previous studies (Alsehibany & Abdelhalim, 2023; Chan & Liou, 2005; Gilmore, 2009; Kaur & Hegelheimer, 2005; Kazaz, 2015; Oktavianti et al., 2022; Sripicharn, 2003; Sun & Wang, 2003; Yoon & Hirvela, 2004; Youssef, 2021; Varley, 2009), which explored students' perceptions of DDL. The deeper analysis appears to have revealed that the learners showed enthusiasm towards using the concordance program and found DDL fun mostly because it was computer-based. Given that the learners are highly dependent on technology, and spend a great deal of their time on computers, it seems reasonable that they opt for learning English through technology and hands-on experience. This finding is also consistent with Chao[2], who pointed out the connection between the positive attitude and the technology-assisted nature of DDL. The results highlight the importance of integrating computer-based corpus analysis in classrooms rather than paper-based concordance as learners show enthusiasm towards DDL practices especially because they are computer-aided.

Another significant finding to emerge from the analysis is that the learners do not find studying vocabulary through DDL more difficult than coursebooks. This is contrary to the previous study by Vannestal & Lindquist (2007), who reported negative feelings of especially weak students towards corpus use. The difficulty of practicing DDL is usually associated with the process of using a concordance program, its display, and the language level of the concordance output (Chatpunnarangsee, 2015). However, in the present study, learners received a demo session, and studied through a

[2] Chao, P. (2010). *A study of collocation learning of junior high students in Taiwan via concordance.* [Paper presentation]. International conference on English teaching (2010), Kaohsiung, Taiwan. http://www2.kuas.edu.tw/edu/afl/20100430Final/Word/2010comp_EPCA.pdf.

pedagogical corpus compiled through texts of elementary and intermediate materials, which eased the burden on the shoulder of the learners. Employing a specialized corpus likely increased the learners' ability to make sense of the data and reduced the need for teacher supervision and monitoring (Aston, 2001). Hence, no feedback was received concerning the challenges of corpus analysis after the introductory session. This finding has significant practical implications for the integration process of DDL. Accordingly, guidance is needed at the beginning stages to help students familiarize themselves with both the concordance program and the concordance output. Classroom handouts play a crucial role in guiding students in the process of dealing with language if teachers have a predetermined aim in mind, such as introducing certain vocabulary items, grammar subject, or lexico-grammatical structures. At this point, there are implications for material developers as well. Since developing such materials is time-consuming for teachers, they need to design more DDL activities and corpus-based materials to motivate teachers in utilizing those ready-made resources without much effort.

In company with general contentedness, the students also perceived concordance-based practice and activities helpful for several reasons. First, they reported significant gains in word forms and word usage. Interview results also revealed that DDL helped students become familiar with lexico-grammatical patterns and collocations which were not intended to be introduced in the first place probably because DDL stands out in revealing grammatical and lexical structures by making patterns more visible (McEnery & Xiao, 2011). This supports Varley's (2009) findings regarding the positive effect of DDL in improving awareness of collocational expressions and lexico-grammatical patterns.

Another notable finding of the study was on the preferences of learners on which type of instruction they would like to receive for vocabulary learning. Although all students expressed positive views about DDL, they all recommended it as a supplementary study rather than a substitute for textbook instruction. The reason behind this preference could be the learning habits of the learners, who received conventional instruction during their primary school years. Naturally, learners feel safe when the information is presented to them, but feel challenged when they are required to use their cognitive skills while dealing with corpus output (O'Sullivan, 2007). Following the students' suggestions on integrating DDL into conventional teaching might create an effective learning environment especially for learners with different learning styles. In line with students' preferences, Meunier (2002) also suggests using corpora as a complementary method as not all types of exercises are compatible with corpus practice.

The present study also points to the need for teacher training. Although the significance and the effect of corpora on language teaching have been widely recognized, teaching programs on the use and evaluation of corpus materials are undervalued (McCarthy, 2008). Indicating that teachers have been "consumers" of corpus materials, McCarthy (2008) maintains that consumer teachers can turn into "active" users of corpus-based materials when corpora are integrated into teacher education; teachers are trained on the evaluation and use of corpus materials; corpora are built when necessary; teachers become "central stakeholders" or "lobbyists", and gain recognition in language education research (p. 565). At a more practical level, Granath (2009) points out the importance of training teachers on using corpora at the university level in courses such as "syntax, written proficiency and translation", which will get them to adopt the habit of consulting corpus rather than just resorting to dictionaries and grammar books (p. 47). For this to happen, it is important that teachers are equipped with corpus literacy. Ma et al. (2021) highlight that while corpus literacy functions as an educational resource, it does not automatically equip educators with the pedagogical skills needed to effectively integrate corpora into classroom teaching. Their research underscores the necessity for teachers to acquire corpus-based language pedagogy (CBLP), which involves integrating corpus linguistics technology into language pedagogy to enhance language instruction.

CONCLUSION

Our study delved into the effects of corpus and data-driven practices on students' vocabulary learning, contributing to the evolving literature on integrating corpora into EFL instruction. Notably, our findings demonstrated the superiority of Data-Driven Learning (DDL) in expanding and retaining vocabulary when compared to traditional coursebook-based instruction. Learner perceptions also yielded positive attitudes toward DDL and concordancing, emphasizing the tangible benefits of these approaches.

This study underscores the significant potential of pedagogical corpora, particularly those sourced from textbooks and instructional materials, in facilitating vocabulary learning, especially for low-level learners. These corpora offer direct access without the complexity of advanced sentence structures, compensating for the limitations of traditional coursebooks in presenting words with diverse meanings and forms, yet incorporating corpora into the classroom necessitates careful planning, teacher guidance, and instructional material support.

Our findings have broader implications for language education and invite educators to consider the potential benefits of corpus-driven practices. As we look ahead, the exploration of different corpus types and their roles in supporting low-level learners remains an important question, offering opportunities for future studies to continue advancing vocabulary instruction in diverse contexts. In future research, we encourage a comprehensive exploration of DDL's impact

on productive vocabulary skills and a systematic analysis of various corpus types. We believe that such investigations will further enhance our understanding of vocabulary acquisition.

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHOR CONTRIBUTIONS

Sibel Tosun: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Software; Validation; Writing – original draft; Writing – review & editing.

Hatice Sofu: Conceptualization; Investigation; Methodology; Resources; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing.

REFERENCES

Alsehibany, R. A., & Abdelhalim, S. M. (2023). Enhancing academic writing vocabulary use through direct corpus consultation: Saudi English majors' perceptions and experiences. *Arab World English Journal*, *14*(1) 166-182. https://dx.doi.org/10.24093/awej/vol14no1.11

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

Balcı, Ö., & Çakır, A. (2011). Teaching vocabulary through collocations in EFL Classes: The case of Turkey. *International Journal of Research Studies in Language Learning, 1*(1), 21-32. https://doi.org/10.5861/ijrsll.2012.v1i1.31

Balunda, S.A. (2009). *Teaching academic vocabulary with corpora: Student perceptions of data-driven learning* [Master's thesis]. Indiana University. https://scholarworks.iupui.edu/handle/1805/2049

Barabadi, E., & Khajavi, Y. (2017). The effect of data-driven approach to teaching vocabulary on Iranian students' learning of English vocabulary. *Cogent Education*, *4*(1), 1283876. https://doi.org/10.1080/2331186X.2017.1283876

Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. University of Michigan Press. https://doi.org/10.3998/mpub.371534

Bergström, D., Norberg, C., & Nordlund, M. (2023). Do textbooks support incidental vocabulary learning? –a corpus-based study of Swedish intermediate EFL materials. *Education Inquiry*, 1-19. https://doi.org/10.1080/20004508.2022.2163050

Binkai, J. (2012). An empirical study on corpus-driven English vocabulary learning in China. *English Language Teaching*, *5*(4), 131–137. https://doi.org/10.5539/elt.v5n4p131

Boulton, A. (2008). Looking for empirical evidence of data-driven learning at lower levels. In B. Lewandowska-Tomaszczyk (Ed.), *Corpus linguistics, computer tools, and applications: State of the art* (pp. 581-598). Peter Lang.

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

Boulton, A. (2010). Data-Driven learning: Taking the computer out of the equation. *Language Learning, 60*(3), 534-572. https://doi.org/10.1111/j.1467-9922.2010.00566.x

Boulton, A. (2012) Language awareness and medium-term benefits of corpus consultation. In A. Gimeno Sanz (Ed.), *New trends in CALL – working together* (pp. 39–46). Macmillan.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348-393. https://doi.org/10.1111/lang.12224

Braun, S. (2005) From pedagogically relevant corpora to authentic language learning contents. *ReCALL 17*(1), 47-64. https://doi.org/10.1017/s0958344005000510

Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching*, *52*(4), 460-475. https://doi.org/10.1017/S0261444819000089

Chan, T., & Liou, H. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning, 18(3),* 231-250. https://doi.org/10.1080/09588220500185769

Chatpunnarangsee, K. (2015). Corpora, concordancing and collocations. *Journal of English Studies, 10*, 102-136.

Chen, W. A. N. G., & Yuhua, L. I. U. (2023). Measuring vocabulary use in Chinese tertiary textbooks: Potentials for incidental vocabulary learning. *Asian Journal of English Language Teaching, 32*(1), 131-154.

Chen, Y.H. (2004). The use of corpora in the vocabulary classroom. *The Internet TESL Journal 10*(9). http://iteslj.org/Techniques/Chen-Corpora.html

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

Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly, 34*(3), 548–60. https://doi.org/10.2307/3587743

Cook, T. D., & Campbell, D. T. (1979). *Quasi-Experimentation: Design and analysis issues for field settings*. Houghton Mifflin.

Creswell, J. W. (2012). *Educational Research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson Education.

El-Esery, A. M. (2015). Corpus concordance for enhancing EFL learners' vocabulary retention. *Journal of Education and Training*, *3*(1), 56-64. https://doi.org/10.5296/jet.v3i1.8202

Flowerdew, L. (2012). *Corpora and language education.* Palgrave Macmillan. https://doi.org/10.1057/9780230355569

Friginal, E., & Roberts, J. (2022). Corpora for materials design. In *The Routledge Handbook of corpora and English language teaching and learning* (pp. 131-146). Routledge. https://doi.org/10.4324/9781003002901-11

Gilmore, A. (2009). Using online corpora to develop students' writing skills. *ELT Journal, 63*(4), 363-372. https://doi.org/10.1093/elt/ccn056

Gilquin, G. (2021). Using corpora to foster L2 construction learning: A data-driven learning experiment. *International Journal of Applied Linguistics, 31*(2), 229-247. https://doi.org/10.1111/ijal.12317

Girgin, U. (2011). *Corpus-based actıvıtıes at lower levels of EFL proficiency: The effectiveness of usıng concordance lınes on grammar learnıng* [Master's thesis]. Bilkent University. http://repository.bilkent.edu.tr/handle/11693/15191

Golabi, N. (2022). The effect of concordancing based teaching on student's vocabulary retention. *Journal of Research in Techno-based Language Education, 2*(1), 25-32.

Gouws, R. (2021). Expanding the use of corpora in the lexicographic process of online dictionaries. In M. Piosik, J. Taborek, & M. Woźnicka (Eds.), *Korpora in der Lexikographie und Phraseologie: Stand und Perspektiven* (pp. 1-20). De Gruyter. https://doi.org/10.1515/9783110716955-001

Granath, S. (2009). Who benefits from learning how to use corpora? In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 47–65). John Benjamins. https://doi.org/10.1075/scl.33.07gra

Harmer, J. (1993). *The practise of English language teaching*. Longman.

Huang, Z. (2014). The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL*, *26*(02), 163-183. https://doi.org/10.1017/s0958344014000020

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press. https://doi.org/10.1017/cbo9781139524773

Jablonkai, R. R., & Čebron, N. (2017). Corpora as tools for self-driven learning: A corpus-based ESP course. In N. A. Alias & J. E. Luaran (Eds.), *Student-Driven learning strategies for the 21st century classroom* (pp. 274-298). IGI Global. https://doi.org/10.4018/978-1-5225-1689-7.ch018

Johns, T. (1991). Should you be persuaded two samples of data-driven learning materials. In T. Johns & P. King (Eds.), *Classroom concordancing (ELR Journal 4)* (pp. 1-16). Centre for English Language Studies.

Karras, J. N. (2016). The effects of data-driven learning upon vocabulary acquisition for secondary international school students in Vietnam. *ReCALL, 28*(2), 166-186. https://doi.org/10.1017/S0958344015000154

Kaur, J., & Hegelheimer, V. (2005). ESL students' use of concordance in the transfer of academic word knowledge: An exploratory study. *Computer Assisted Language Learning*, *18*(4), 287–310. https://doi.org/10.1080/09588220500280412

Kavanagh, B. (2021). Norwegian in-service teachers' perspectives on language corpora in teaching English. *Nordic Journal of Language Teaching and Learning, 9*(2), 90-106. https://doi.org/10.46364/njltl.v9i2.933

Kazaz, I. (2015). *Corpus-aided language pedagogy: The use of concordance lines in vocabulary instruction* [Master's thesis]. Bilkent University. http://repository.bilkent.edu.tr/handle/11693/16944

Koosha, M., & Jafarpour, A. (2006). Data-driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners. *Asian EFL Journal Quarterly, 8*(4), 192-209.

Lee, C. Y., & Liou, H. C. (2003). A study of using web concordancing for English vocabulary learning in a Taiwanese high school context. *English Teaching and Learning, 27*(3), 35-56.

Lee, H., Warschauer, M. & Lee, J. H. (2019) The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics, 40*(5), 721-753. https://doi.org/10.1093/applin/amy012

Lee, P., & Lin, H. (2019). The effect of the inductive and deductive data-driven learning (DDL) on vocabulary acquisition and retention. *System*, *81*, 14-25. https://doi.org/10.1016/j.system.2018.12.011.

Ma, Q., Tang, J., & Lin, S. (2022). The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online collaboration. Computer *Assisted Language Learning, 35*(9), 2731-2760. https://doi.org/10.1080/09588221.2021.1895225

McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching, 41*(4), 563-574. https://doi.org/10.1017/s0261444808005247

McEnery, T., & Xiao, R. (2011). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (vol. 2, pp. 364-380). Routledge. https://doi.org/10.4324/9780203836507

Melissourgou, M. N. & Frantzi, K. T. (2019). Stories, second language writers, and genre awareness. In V. Kourtis-Kazoulis, T. Aravossitas, E. Skourtou, & P.P. Trifonas (Eds.), *Interdisciplinary Research Approaches to Multilingual Education* (pp. 117-129). Routledge

Meunier, F. (2002). The pedagogical value of native and learner corpora in EFL grammar teaching. In Granger, S., Hung, J. & Petch-Tyson, S. (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 119-141). John Benjamins. https://doi.org/10.1075/lllt.6.10meu

Meunier, F. (2011). Corpus linguistics and second/foreign language learning: Exploring multiple paths. *Revista Brasileira de Linguística Aplicada*, *11*(2), 459-477. https://doi.org/10.1590/S1984-63982011000200008

Meunier, F., & Gouverneur, C. (2009). New types of corpora for new educational challenges: Collecting, annotating and exploiting a corpus of textbook material. In K. Aijmer (Ed.), *Studies in corpus linguistics* (vol. 33, pp. 179–201). John Benjamins. https://doi.org/10.1075/scl.33.16meu

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to classroom. language use and language teaching*. Cambridge University Press. https://doi.org/10.1017/cbo9780511497650.009

O'Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL, 19*(3), 269-286. https://doi.org/10.1017/s095834400700033x

Oktavianti, I. N., Triyoga, A., & Prayogi, I. (2022). Corpus for language teaching: Student' perceptions and difficulties*. PROJECT (Professional Journal of English Education), 5*(2), 441-455. https://doi.org/10.22460/project.v5i2.p441-455

Pérez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011–2015. *Computer Assisted Language Learning, 35*(1-2), 36-61. https://doi.org/10.1080/09588221.2019.1667832

Römer, U. (2008). Corpora and language teaching. In A. Ludeling & M. Kyto (Eds.), *Corpus linguistics: An international handbook* (vol. 1, pp. 112-130). Mouton de Gruyter.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language teaching research*, *12*(3), 329-363. https://doi.org/10.1177%2F1362168808089921

Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly, 36*(2), 145-171. https://doi.org/10.2307/3588328

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education.* Benjamins. https://doi.org/10.1075/scl.22

Sinha, T. S. (2021). EFL learners' perception of and attitude to corpus as a vocabulary learning tool. *The Reading Matrix: An International Online Journal, 21*(2), 106-119.

Sripicharn, P. (2003). Evaluating classroom concordancing: The use of concordance-based materials by a group of Thai students. *Thammasat Review, 8*(1), 203–236.

Sun, Y. C., & Wang, L. Y. (2003). Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning*, *16*(1), 83-94. https://doi/10.1076/call.16.1.83.15528

Sun, Y., & Dang, T. N. Y. (2020). Vocabulary in high-school EFL textbooks: Texts and learner knowledge. *System, 93*, 102279. https://doi.org/10.1016/j.system.2020.102279

Toriida, M. C. (2016). Steps for creating a specialized corpus and developing an annotated frequency-based vocabulary list. *TESL Canada Journal*, *34*(1), 87-105. https://doi.org/10.18806/tesl.v34i1.1257

Vannestal, M. E., & Lindquist, H. (2007). Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course*. ReCALL, 19*(3), 329-350. https://doi.org/10.1017/S0958344007000638

Varley, S. (2009). I'll just look that up in the concordance: Integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning, 22*(2), 133-152. https://doi.org/10.1080/09588220902778294

Widdowson, H.G. (2003). Expert beyond experience: Notes on the appropriate use of theory in practice. In. Newby, D. (Ed.), *Mediating between theory and practice in the context of different learning cultures and languages* (pp. 23–30). Council of Europe.

Wu, S., Witten, I.H. & Franken, M. (2010). Utilizing lexical data from a Web-derived corpus to expand productive collocation knowledge. *ReCALL, 22(1),* 83-102. https://doi.org/10.1017/s0958344009990218

Yoon, H., & Hirvela, A. (2004). ESL 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

Youssef, A. F. F. A. (2021). Learners' Attitudes Towards Utilizing Corpus-based Approach in the Saudi EFL *Context. College of Arts Journal, 5*(58), 5-36. https://doi.org/10.21608/jfabsu.2021.159373

APPENDIX A

Sources used for the corpus

Acevedo, A., Lethaby, C., & Harmer, J. (2007). *Just listening and speaking: For class or self study*. Marshall Cavendish Education.

BBC (2007). *BBC learning English: Talk about English webcast: Archive.* http://www.bbc.co.uk/worldservice/learningenglish/webcast/tae_insight_archive.shtml

Bernard, L. (2002). *Absolute English*. http://absolutenglish-972.pagesperso-orange.fr/notes/carnival/raw/_starthere.htm

Blanchard, K. & Bonesteel, L. (*2010). From reading to writing 2*. Pearson Education.

British Council (n.d). *Learn English teens: listening*. http://learnenglishteens.britishcouncil.org/skills/listening-skills-practice

Dummett, P., Hughes, J., & Stephenson, H. (2015). *Life: Student book 2*. Cengage Learning.

Evans, V. & Dooley, J. (2000). *Reading & writing targets 2*. Express Publishing.

Evans, V. & Dooley, J. (2000). *Reading & writing targets 3*. Express Publishing.

Greet, J. (Ed.). (2010). *Timesaver intermediate listening.* Mary Glasgow Magazines.

Hill, L. A. (1998). *Elementary stories for reproduction*. Oxford University Press.

Hill, L. A. (1998). *Intermediate stories for reproduction*. Oxford University Press.

Kandiller, B., & Velioğlu, A. (2012). *Reader at work 1*. Ankara: METU Press.

MacIntyre, T. (2009*). Reading explorer 2*. Boston: Heinle Cengage Learning.

Mc Carthy, M., Mc Carten, J., & Sandiford, H. (2005*). Touchstone level 2, student's book.* Cambridge University Press.

Ostrowska, S. (2014). *Unlock 2: Reading & writing skills*. Cambridge University Press.

Oxenden, C., Latham-Koenig, C., & Seligson, P. (2009). *New English file: Pre-intermediate student's book*. Oxford University Press.

Oxenden, C., Latham-Koenig, C., & Seligson, P. (2004). *New English file: Intermediate Student's book*. Oxford University Press.

Publishing, S. (2002). *Reading comprehension skills and strategies level 3*. Saddleback Educational Publishing.

Richards, J. C., & Sandy, C. (2008). *Passages* (2nd edition). Cambridge University press.

Scott, R. (1987). *Reading: Elementary*. Oxford University Press

Türkan, C. C., & Türkan, N. (2006). *More to read 1*. METU Press.

APPENDIX B

Sample Concordance Handout

Analyse the words/phrases given and answer the following questions.

1. Search for join*

What is the part of speech of the word "join"? _____

What preposition(s) directly follows "*join*"? _____

List 6 phrases that collocate with *the word "join"* as an object of the verb. (e.g.: join a racing school) _____

2. Search for appl*

Identify 2 different parts of speech in the concordance lines (e.g.: noun, verb, adjective, adverb, etc.) and write 2 example concordance lines below. _____

Find a concordance line in which "apply" is used in an infinitive construction. Write it below. _____

What verbs precede the infinitive form? _____

What preposition(s) directly follows *apply*? _____

List the words or phrases after the preposition(s) you wrote above. _____

3. Search for look* after

What is the part of speech of "look after"? _____

Find a concordance line in which "look after" is used as a main verb. Write it below. _____

Find a concordance line in which "look after" is used in an infinitive construction. Write it below. _____

List 6 phrases that collocate with *"look after"* as the object of a verb. (e.g.: look after my family) _____

What do the sentences have in common? What do people normally 'look after'? _____

Looking at the concordance lines, write a sentence using "look after". _____