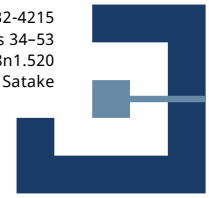
Regular Paper

The JALT CALL Journal ISSN 1832-4215
Vol. 18, No.1 Pages 34–53
https://doi.org/10.29140/jaltcall.v18n1.520



Castledown

The effects of corpus use on L2 collocation learning

Yoshiho Satake
t31330@aoyamagakuin.jp
Aoyama Gakuin University, Japan

In data-driven learning (DDL) settings, few studies have examined the specific effects of corpus use on language learning, especially in the study of collocations. This study examined the effects of corpus use on L2 collocation learning. It remains unclear how corpora can contribute to learning collocations and what types of information learners can access and process from these corpora. This study aimed to investigate the effects of corpus use on learning L2 collocations, specifically how information extracted from corpus data assisted learners in retaining and producing collocations. The effects of corpus use were examined by comparison with the effects of dictionary use. Fifty-five intermediate English as a foreign language (EFL) learners from a university in Tokyo looked up verb-noun collocations by consulting either the corpus or the dictionaries, noted related information, and wrote short essays on the topics related to the collocations. Pre- and post-tests were conducted to assess learning. It was found that corpus use promoted a greater collocation output than dictionary use because of the many example sentences and frequency of co-occurrence words. The corpus users focused on word combinations and looked up high-frequency collocations, while the dictionary users focused on meaning, then looked up and memorized the collocations. Findings suggest that teachers need to consider the information that learners access and process when using corpora and the specific effects of corpus use on L2 learning for more effective vocabulary teaching in DDL settings.

Keywords: data-driven learning (DDL), corpus use, second language acquisition (SLA), English language learning (ELL), vocabulary learning

Introduction

The size and complexity of corpora have been growing to include various tools that enable automatic identification of different linguistic features and profiles, such as collocations, grammatical patterns, keywords, and stylistic

profiles (Leńko-Szymańska & Boulton, 2015). Corpora have influenced research in various fields, including language pedagogy.

In the application of corpus linguistics in language teaching, there are two forms of corpus use: direct and indirect. Indirect uses include compiling dictionaries and teaching materials based on corpus information, such as word frequency. Learners do not directly use a corpus, but rather products based on corpus information. On the other hand, direct corpus use involves learners directly accessing and analyzing corpus information to help them learn the language. Johns (1991), one of the pioneers of direct corpus use in language learning, created the term data-driven learning (DDL) for this approach. Leech (1997) states that a data-driven approach to learning should hold an important position in corpus use for pedagogical purposes because language learning is based on usage. According to Flowerdew (2015), DDL is supported by certain learning theories, such as the noticing hypothesis, constructivist learning, and Vygotskyan sociocultural theories.

Although it was expected in the 1990s that learners would directly access corpus information and become autonomous DDL learners when corpus tools became more powerful (Sinclair, 1997), DDL is not widely used in second language (L2) teaching at school (e.g., Hirata, 2019; Vyatkina, 2020).

According to Leńko-Szymańska and Boulton (2015), there are hurdles to introducing DDL in class. Teachers think DDL is a difficult teaching technique and are reluctant to introduce it into their classrooms because learners need to have sufficient skills to use and analyze corpus information. Often, low-proficiency learners have difficulty using corpora of target languages because of their limited language knowledge (e.g., Chujo *et al.*, 2015; Mizumoto *et al.*, 2016; Mueller & Jacobsen, 2016). In addition, communicative language teaching is another hurdle to DDL because communicative teaching focuses on fluency more than accuracy, while DDL focuses on accuracy more than fluency (Widdowson, 1990). Moreover, Leńko-Szymańska and Boulton (2015) state that the most significant hurdle is that DDL is considered an inductive approach to learning (Flowerdew, 2015) requiring learner autonomy to analyze and interpret corpus information individually, and such autonomy is not necessarily compatible with school traditions or learners' learning styles (Boulton, 2009a). Teachers often complain about a lack of instruction on how to make the most of a corpus (Leńko-Szymańska, 2014), and their perception of and ability in corpus use could be another hurdle (Leńko-Szymańska & Boulton, 2015).

Although studies show that these hurdles can be cleared (e.g., Boulton, 2015; Chujo *et al.*, 2015, 2016; Hirata, 2019; Saeedakhtar, 2020), DDL is not well integrated into L2 teaching at school. Compared to the use of dictionaries, which are traditional reference materials, DDL is considered to require considerable training and is time-consuming (Boulton & Cobb, 2017), and there are not enough studies to show the superiority of using corpora over dictionaries. More longitudinal studies are needed to prove the effectiveness of DDL in L2 language learning, convince L2 language teachers of DDL's strengths, and lead them to use a corpus in the classroom. This study deals with the problem of how and where the use of a corpus could best help in learning L2 vocabulary.

In particular, the author focuses on how L2 collocations can be memorized and transformed into output by accessing the information in a corpus. This study hypothesized that being exposed to many sample sentences using the target word(s) through corpus use helps memorize and output the collocations of these words because learners can induce the appropriate pattern from the examples. To test the hypothesis, pre- and post-tests were conducted before and after the vocabulary learning and essay writing tasks to examine how and to what extent corpus evidence helped learners learn L2 collocations.

In this paper, the author begins with a review of empirical research on the effects of DDL in L2 vocabulary learning. This is followed by the research questions and description of the experimental design. Finally, the author presents the results of the experiment and discusses the theoretical, methodological, and pedagogical implications of the study. The author expects that this study can enhance second language learning and teaching approaches by providing insights into the effectiveness of DDL for L2 learning.

Literature review

Since the Brown Corpus – the first computer-based corpus with approximately one million words – was constructed in 1964, many corpora have been compiled. As for pedagogical use, corpora were first used indirectly in making instructional materials for language teaching and learning, such as dictionaries and word lists. They were later used directly for learning vocabulary and grammar. This method of learning from corpus data is called DDL. DDL is a way of learning in which learners are exposed to an enormous quantity of authentic data from a corpus and encouraged to autonomously investigate the language and infer its patterns in it (Boulton, 2009b).

DDL is supported by the noticing hypothesis (Flowerdew, 2015), which argues that the acquisition of linguistic input is enhanced when learners pay attention to linguistic features (Schmidt, 1993, 2001). Frequency information helps learners notice the form of the target language (Swain, 1998). Given that corpora provide information about the frequency with which learners notice the form of the target language and process the language, we can say that the noticing hypothesis is realized in DDL.

Meta-analyses of DDL have recently shown that DDL is effective in language learning (Boulton & Cobb, 2017; Lee *et al.*, 2019; Mizumoto & Chujo, 2015), and Boulton (2009c) showed the effects of corpus use. He divided 132 intermediate and lower-level learners of English into four groups and compared how they dealt with linking adverbials (such as “but” and “in fact”) through the use of traditional resources (such as dictionaries or grammar books) or corpus use, finding that corpus use, especially concordance lines, contributed to answering fill-in-the-blank questions more than traditional resources. Boulton (2008) also suggested that consulting printed concordance lines helped lower-intermediate learners answer binary choice questions accurately.

As for the effects of DDL on memorizing and outputting collocations, Saeedakhtar *et al.* (2020) compared the effects of hands-on DDL (learners use

corpora directly) and hands-off DDL (learners use teacher-made concordance data in handouts) on learning verb-preposition collocations. Sixty Iranian pre-intermediate learners of English were divided into three groups: hands-on, hands-off, and control. During 10 treatments, the hands-on DDL group accessed the corpus information to search 66 target collocations, while the hands-off DDL group accessed paper-based concordance lines. The control group explicitly learned the same collocations using traditional approaches. The results show that the DDL groups outperformed the control group in memorizing the target collocations on the immediate post-test, and the hands-on group outperformed the hands-off group on the delayed post-test. The findings suggest that both hands-on and hands-off DDL help pre-intermediate learners improve their collocational knowledge. Satake (2020b) also examined the effects of corpus use on learning L2 collocations. Sixty Japanese intermediate learners of English were divided into an experimental group (corpus use) and a control group (dictionary use), and the collocations of the target words were searched to learn collocational knowledge. The results show that the experimental group outperformed the control group in promoting word associations and output of collocations, while both groups effectively memorized collocations. The findings suggest that corpus consultation improves learners' understanding of usage.

From the above studies, DDL seems to be an effective approach for L2 vocabulary learning, such as memorizing linking adverbials and memorizing and outputting collocations. Contrary to language teachers' typical assumption, DDL works well for intermediate-and lower-level English learners as well as advanced learners, as Johns (1991) argued. Aside from direct corpus use, consulting printed concordance lines also helps.

However, various issues remain unclear. According to Tono (2015), scholars have not thoroughly discussed the relation between the kind(s) of information that learners access, process, and acquire with corpus use and their computer skills. Although Flowerdew (2010) states that corpora help learners acquire phraseological patterning such as collocations, because collocations are not easily accessible in dictionaries, the specific effects of corpus use on collocation learning remain unclear. Considering that "language knowledge is collocational knowledge" (Nation, 2001, p. 318) and thus that understanding collocations is an important part of language learning, more empirical research is needed to judge whether corpus use is effective in improving L2 collocational knowledge. Since corpora are not used much in L2 classrooms in Japan, and the number of long-term studies (e.g., Satake, 2020a; Chujo *et al.*, 2015; Hadley & Charles, 2017) is limited, more studies in Japanese settings are needed to show the specific effects of corpus use on L2 learning so that more Japanese teachers might be convinced of the effectiveness of DDL and use it in their classes. Therefore, this study examined how corpus information such as frequency helps learners notice, memorize, and output appropriate collocations.

Method

Aims and research questions

This study investigated the effects of corpus use on memorizing and outputting collocations, compared to the effects of dictionary use. The following research questions were addressed:

1. Do corpora and dictionaries have different effects on memorizing collocations?
2. Do corpora and dictionaries have different effects on outputting of collocations?
3. Do corpus users and dictionary users access and process different information?

Participants

The 55 participants were divided into two groups: the experimental group (28 corpus users) and the control group (27 dictionary users). They were intermediate Japanese learners of English, who were university sophomores majoring in English and American literature at a private university in Tokyo. The students were around 20 years old, and the ratio of male to female students was one to four, with the majority being female. They had studied English for seven years in school – six years in junior high school and high school, and one year in university. The majority of the students had passed the Jitsuyo Eigo Gino Kentei (EIKEN) Test in Practical English Proficiency Grade 2, which is roughly equivalent to B1 in the Common European Framework of Reference for Languages (CEFR). The participants took the compulsory English reading course taught by the author. Before starting the study, the author explained the study in Japanese and reassured them that their anonymity would be preserved, that they had the right not to participate in it, and that their participation or non-participation would not affect their grade. The author asked the students for permission to use their data, obtaining written permission from 55 out of 60 students.

Instruments

The corpus. The British National Corpus (BNC) was used for this study because it is a large, balanced corpus that functions as a normative model for learners of English. The BNC is an approximately 100-million-word corpus of written and spoken British English from the latter half of the 20th century. To enable students to use the BNC easily, this study used a publicly available online corpus query system called IntelliText by the Centre for Translation Studies at the University of Leeds (Wilson *et al.,* 2010). With IntelliText, users can search various corpora, including the BNC. It has a user-friendly interface for users to search for words and phrases. Before the tasks, the students who used the BNC were provided with 20 minutes of instruction on how to use the corpus tools. The author provided learners with instructions on how to search for

high-frequency collocations with verbs that come up to three words before the target nouns, and how to sort and interpret concordance lines.

Dictionaries. The author allowed the students to use any dictionaries they liked because imposing only one dictionary for all the students was impractical, considering the many kinds of dictionaries they had. The majority of the students used English-Japanese dictionaries, which is reasonable, as foreign language learners tend to use bilingual dictionaries irrespective of their language proficiency levels (Piotrowski, 1989). Approximately half of the students used the online Weblio English-Japanese Dictionary, which contains more than ten million words (2018). All the students except one used online dictionaries.

Tasks. Two types of timed tasks were used in this study, and both were timed for 10 minutes. Before the first task, special instructions were provided to introduce the use of concordancers as a reference resource. The target collocation for Week 1 was "verb + law" and the collocation for Week 2 was "verb + pregnancy." The words "law" and "pregnancy" were chosen as the target words because they were often used in articles the students were reading in class. One article was about sales of organs and laws, and the other was about whether it is ethical to plan pregnancy to have a baby whose marrow would be a close match for a leukemia patient (Smith & Mare, 2011). Although there are not many collocations of "verb + pregnancy," "pregnancy" was chosen as the target item to deepen the students' vocabulary knowledge because there was a gap between their impression of the word and how it was used in the article. Many students said that they thought "pregnancy" had a positive connotation such as "congratulations" or "happiness," while the word was used with verbs with a negative connotation such as "terminate" and "end."

In the first task, after searching the textbook for verbs and the target noun collocations, the students were asked to look up three or more verb-noun collocations of the target words. A BNC was used by the experimental group. They searched for verb and target noun collocations and read concordance lines. The control group used dictionaries for their choices. Students typed collocations they looked up and what they found using Microsoft Word so that the author could collect their data in Word files and analyze their searches and findings later. For the second task, the participants were asked to write a short essay on the topic related to each target noun. They shared the same topics. For example, the topic for "pregnancy" was whether "it is acceptable for parents to conceive a child to provide an organ or tissue that will save the life of another one of their children" (Smith & Mare, 2011, p. 177). The author expected that students would use the collocations they looked up.

Pre- and post-tests. Pre- and post-tests were administered in this study and timed for five minutes each. Before the tasks, the author provided students with a pre-test for each target noun, which had four fill-in-the-blank questions with Japanese translation. There was a blank before the target noun, and the

students were supposed to fill in each blank with an appropriate verb. The following is an example of these questions.

A new computerized nationwide registry will be used to help ( ) the law.

全国的新電子登録がその法律を施行するのを助けるために使用されるだろう。

Out of the four verb-noun collocations in the questions, the two verb-noun collocations were used in the textbook, and the other two verb-noun collocations were two of the high-frequency verb and the target noun collocations in the BNC. The author selected comparatively high frequent verb-noun collocations for the questions so that students could easily find the verb-noun collocations in the BNC or dictionaries when they looked them up later. One point was given for each correct answer; thus, the full score was 4. Two weeks after the tasks, the author gave the students the post-test. It was almost the same as the pre-test, except that the post-test also asked the students to write phrases using as many target nouns as possible. A collocation was regarded as appropriate if it appeared in the BNC.

Procedure

The following procedure was adopted:

1. The pre-test was conducted (5 minutes).
2. Instruction was provided for IntelliText (20 minutes).
3. Task sessions were held. The timed task for finding verb-noun collocations and the timed essay task using the target collocations were given (10 minutes for each task, 20 minutes total). This task session was repeated the following week using different target words.
4. Two weeks after the second task session, the post-test was conducted (5 minutes).

The two-group (corpus users vs. dictionary users) pre-post design was used to investigate the effects of corpus use on learning collocations. The two classes of students were given five minutes to take the pre-test described above. Before the tasks, the students of the experimental group were given 20 minutes of instruction on how to use the corpus tools: they were instructed to search for verb-noun collocations and interpret the concordance lines. Then, the two classes of students were given 20 minutes to perform the two tasks described above. The students in the experimental group consulted the BNC to complete the tasks, while the students in the control group consulted dictionaries of their choice. The next week, the task session was repeated for the different target words. Two weeks after the task stage, the students were given five minutes to take the post-test, which was almost the same as the pre-test. The author compared the results from the pre- and post-tests to examine whether the tasks were effective for memorizing and outputting the target collocations. The author also compared the number of collocations that the students consulted

with corpus use and that with dictionary use to investigate how students learned collocational knowledge of words through a corpus or dictionaries.

Results

Effects of corpus use on memorizing collocations

Table 1 summarizes the average marks for fill-in-the-blank questions for each target noun in the pre- and post-tests, categorized by reference materials.

Table 1. The average scores on the fill-in-the-blank questions

	Experimental group (Corpus users; *n* = 28)		**Control group (Dictionary users; *n* = 27)**	
	law	pregnancy	law	pregnancy
Pre-test	2.39	2.21	2.41	2.19
Post-test	2.61	2.71	2.81	2.81

To judge whether there was a significant difference in the average scores between the pre- and post-tests, the author used analysis of variance (ANOVA), which found a significant difference between them for both target nouns (for "law," F (1, 53) = 9.02, p < .01, η_p^2 = .15; for "pregnancy," F (1, 53) = 36.75, p < .01, η_p^2 = .41). However, there was no significant difference in the average marks between corpus users and dictionary users (for "law," F (1, 53) = 0.74, p = .39, η_p^2 = .01; for "pregnancy," F (1, 53) = 0.07, p = .79, η_p^2 = .00) although the control group outperformed the experimental group on both words in the post-test. Thus, both corpus use and dictionary use helped learners memorize the target collocations, even though there was no significant difference between the effects of corpus and dictionary use on memorizing them.

Effects of corpus use on outputting collocations

Output in the post-test. Table 2 summarizes the average number of the verb-noun collocations the students output in the post-test.

Table 2. The average number of the output of the target collocations in the post-test

Experimental group (Corpus users; *n* = 28)		**Control group (Dictionary users; *n* = 27)**	
law	pregnancy	law	pregnancy
0.93	1.71	0.44	1.15

In the post-test, students were told to write verb-noun collocations of the target nouns, "law" and "pregnancy," as many as possible. The author counted

the number of appropriate collocations that they wrote. On average, the BNC users wrote a larger number of appropriate collocations using both target nouns than the dictionary users. To judge whether there was a significant difference in the average number of output of appropriate collocations between the experimental group and the control group, the Mann-Whitney U-test was used because the data did not show a normal distribution. The test found a significant difference between them for both target nouns (for "law," $z = -2.85$, $p < .01$, $r = .38$; for "pregnancy," $z = -2.00$, $p < .05$, $r = .27$); thus, corpus use helped learners output appropriate collocations of the target words in the post-test significantly more than dictionary use did. Since both corpus and dictionary users output more verb- "pregnancy" collocations than verb- "law" collocations, the articles that students were reading might have affected the result. The result suggests that corpus use is more effective than dictionary use in promoting the output of the collocations students looked up. However, since the number of collocations output in the post-test was lower than the average score on the fill-in-the-blank questions for each target noun on the pre-test and post-test, we may say that it was more difficult for students to output collocations than to fill an appropriate verb into each blank. That is, the results suggest that creative use of collocations was more difficult than simply writing collocations.

Output in the essay. Table 3 summarizes the average number of verb-noun collocations the students output in the essay.

Table 3. The average number of the output of the target collocations in the essay

Experimental group (Corpus users; *n* = 28)		**Control group (Dictionary users; *n* = 27)**	
law	pregnancy	law	pregnancy
0.68	0.29	0.37	0.04

After students looked up collocations using the BNC or dictionaries, they wrote a short essay on the topic related to the target nouns. On average, the BNC users used more collocations than dictionary users. To judge whether there was a significant difference in the average number of collocations output between the experimental and control groups, the Mann-Whitney U-test was used because the data did not show a normal distribution. The test found a significant difference between them for both target nouns (for "law," $z = -2.54$, $p < .05$, $r = .34$; for "pregnancy," $z = -2.47$, $p < .05$, $r = .33$), indicating that corpus use helped learners use appropriate collocations of the target words in the essay significantly more than dictionary use. In contrast to the output in the post-test, both corpus and dictionary users used verb- "law" collocations more than verb- "pregnancy" collocations in their essays, perhaps because "law" is an easier word for students than "pregnancy," which could promote more use of verb- "law" collocations than verb- "pregnancy" collocations.

The average number of the collocations that students searched

Table 4 summarizes the average number of target collocations that the students searched with corpus or dictionary use.

Table 4. The average number of the collocations that students consulted

Experimental group (Corpus users; *n* = 28)		**Control group (Dictionary users; *n* = 27)**	
law	pregnancy	law	pregnancy
2.89	3.50	5.30	5.11

For both target nouns, dictionary users searched for more collocations than corpus users. This could be because it was difficult for the learners to search for and interpret corpus information. To judge whether there was a significant difference between the average number of collocations that corpus users and dictionary users searched, the author used the Mann-Whitney U-test because the data did not show a normal distribution. The test found a significant difference between them, for both target nouns (for "law," $z = -6.86$, $p < .01$, $r = .92$; for "pregnancy," $z = -6.72$, $p < .01$, $r = .91$); thus, students searched significantly more words with dictionary use than corpus use. This could be related to the fact that the students had been using dictionaries for years, whereas the corpus was introduced in just 20 minutes. Considering that dictionary users output fewer collocations in the post-test than the BNC users, we may say that the dictionary users did not efficiently use the information they accessed, although they accessed more information. In other words, the BNC users could use the information they accessed more efficiently, although they accessed less information. Since corpus users' fewer searches for collocations suggest that they could spend more time for each collocation than dictionary users, we may say that sufficient time to search for each collocation contributed to more output with corpus use. This suggests that teachers should provide learners with instruction on the appropriate number of corpus searches during the limited time so that learners can effectively use the information they access in the corpus.

There was not only a quantitative difference but also a qualitative difference between the verbs that the corpus users searched and those that the dictionary users searched. Corpus users searched for less diverse verbs than dictionary users. To consider the qualitative difference between the search with corpus use and that with dictionary use, let us see the verbs the students searched for with one of the target nouns, "law." The corpus users' search focused on 17 kinds of verbs, while dictionary users searched 58 kinds of verbs. Table 5 shows the top 10 verbs that the students looked up for verb-law collocations, and Figure 1 shows high-frequency collocations with verbs that come up to three words before "law" in the BNC.

Table 5. The top 10 verbs the students searched for with verb- "law" collocations

	Experimental group (Corpus users; *n* = 28)		Control group (Dictionary users; *n* = 27)	
rank	verb	number of searches	verb	number of searches
1	enforce	22	break	13
2	break	21	obey	10
3	obey	11	be	9
4	pass	7	become	7
5	become	6	enforce	7
6	change	6	lay	7
7	require	5	take	6
8	be	2	violate	6
9	disobey	2	observe	5
10	violate	2	pass	5
			read	5
			study	5

Collocations of law

Collocation	Count	F1	F2	LL	MI	T
law ~~ be	1178	21712	3582022	70.64	0.75	13.95
law ~~ break	303	21712	22888	491.62	6.08	17.15
law ~~ become	212	21712	66107	197.77	4.04	13.67
law ~~ make	121	21712	210492	25.48	1.56	7.27
law ~~ require	114	21712	29240	116.78	4.32	10.14
law ~~ change	109	21712	66365	68.21	3.07	9.2
law ~~ enforce	107	21712	2340	240.02	7.87	10.3
law ~~ pass	87	21712	23031	87.79	4.27	8.85
law ~~ take	86	21712	174068	13.96	1.34	5.61
law ~~ obey	67	21712	1285	154.75	8.06	8.15
law ~~ apply	61	21712	19739	55.75	3.98	7.32
law ~~ study	60	21712	39020	35.78	2.98	6.76
law ~~ lay	54	21712	10338	62.86	4.74	7.07
law ~~ establish	52	21712	17914	46	3.89	6.73
law ~~ exist	48	21712	14780	44.98	4.06	6.51

Figure 1. High-frequency collocations of "verb + law" in the BNC

Corpus users tended to look up more frequent verb- "law" collocations than dictionary users. As Figure 1 shows, when displaying verbs that are found up to three words to the left from "law" in the BNC, the frequent verbs that appear more than 100 times are "be," "break," "become," "make," "require," "change," and "enforce," in the order of high frequency. Table 5 shows that the top 10 verb- "law" collocations corpus users looked up include 6 out of 7 high-frequency verb- "law" collocations in the BNC, while the top 12 verb- "law" collocations dictionary users looked up include only 4 of the 7 high-frequency verb- "law" collocations in the BNC. Overall, in the case of corpus users, 62

out of 91 searches in total, that is, 68.1 percent of their searches, focused on high-frequency collocations in the BNC. However, in the case of dictionary users, 36 out of 133 searches in total, that is, only 27.1 percent of their searches focused on high-frequency collocations. This difference arises from the frequency information of each reference material. Corpus users could access and use the frequency information of the target collocations; thus, they focused on high-frequency collocations. On the other hand, since dictionaries show each usage like catalogs, providing dictionary users little information on the frequency of each collocation, they paid less attention to the frequency of the target collocations than corpus users. Considering that the priority of learning high-frequency collocations is higher than learning low-frequency collocations, the result suggests that corpus use promotes more efficient learning of collocations than dictionary use.

How learners used the information they accessed

Fill-in-the-blank questions. Table 6 summarizes how searching for collocations influenced the answers to the fill-in-the-blank questions in the post-test, showing the relationship between answer accuracy and retrieval/non-retrieval as a percentage (as for the information on the number of searches, see Tables 4 and 5).

Table 6. The effect of search on the answers to the fill-in-the-blank questions

		Experimental group (Corpus users; *n* = 28)		**Control group (Dictionary users; *n* = 27)**	
		law	pregnancy	law	pregnancy
Searched	Correct	30.6%	31.5%	21.4%	19.2%
	Wrong	18.5%	17.4%	8.0%	4.8%
Not searched	Correct	35.2%	33.7%	48.2%	52.9%
	Wrong	15.7%	17.4%	22.3%	23.1%

For both target words, the corpus users' searches promoted more use of the information they accessed than dictionary users' searches. For more than 30 percent of the answers in total, the corpus information was appropriately used and contributed to answering correctly, while the dictionary information contributed to correctly providing only approximately 20 percent of the answers. The results suggest that corpus information was more efficiently used for memorizing collocations than dictionary information. On the other hand, for more than 50 percent of the total answers, neither the corpus information nor the dictionary information was used.

Output in the post-test. Table 7 summarizes the number of output in the post-test and how the search for collocations influenced the output in the post-test.

Table 7. The number of the output in the post-test

	Experimental group (Corpus users; *n* = 28)				Control group (Dictionary users; *n* = 27)			
	law		pregnancy		law		pregnancy	
Searched	12	(46.2%)	22	(45.8%)	1	(8.3%)	9	(29.0%)
Not searched	14	(53.8%)	26	(54.2%)	11	(91.7%)	22	(71.0%)

As in the case of the fill-in-the-blank questions, for both target words, the corpus users' search promoted more use of the information they accessed than dictionary users' search. For approximately 45 percent of the collocations in total, the corpus information was used and contributed to their output, while the dictionary information contributed to outputting less than 30 percent of the collocations. The results suggest that the corpus information was more efficiently used for outputting collocations than the dictionary information. To consider how the students used the information they accessed, let us see the verbs the students used for one of the target nouns, "law." Table 8 shows the verbs of the verb-law collocations that the students searched and wrote in the post-test.

Table 8. The verbs the students searched and used for verb- "law" collocations in the post-test

Experimental group (Corpus users; *n* = 28)		Control group (Dictionary users; *n* = 27)	
verb	number of output	verb	number of output
enforce	3	enforce	1
make	3		
pass	3		
be	1		
become	1		
change	1		

Apart from "make," the verbs that corpus users output in the post-test are included in the top 10 verbs that they searched for verb- "law" collocations, and thus corpus users tended to use frequently searched words in the post-test. In addition, corpus users tended to output frequent verb- "law" collocations, as five out of six verbs ("enforce," "make," "be," "become," "change") are frequent verbs that appear more than 100 times in the BNC. Thus, corpus search promoted more output, especially the output of high-frequency verb- "law" collocations, than dictionary search. On the other hand, for more than 50 percent of the collocations in total, neither the corpus information nor the dictionary information was used, suggesting that more than half of the collocations the students wrote were collocations they already knew.

Output in the essay. Table 9 summarizes the number of the output in the essay and how searching for collocations influenced the output in the essay.

Table 9. The number of the output in the essay

	Experimental group (Corpus users; *n* = 28)				**Control group (Dictionary users; *n* = 27)**			
	law		pregnancy		law		pregnancy	
Searched	7	(36.8%)	2	(25.0%)	2	(20.0%)	0	(0.0%)
Not searched	12	(63.2%)	6	(75.0%)	8	(80.0%)	1	(100.0%)

Similar to the case of the fill-in-the-blank questions and the output in the post-test, for both target words, the corpus users' search promoted more use of the information they accessed than dictionary users' search. For 25.0 to 36.8 percent of collocations in total, the corpus information was used and contributed to their output, while the dictionary information contributed to outputting only 20 percent of collocations using "law" and was not used for the output of collocations using "pregnancy." The results suggest that corpus information was more efficiently used for outputting collocations than dictionary information. To consider how the students used the information they accessed, let us examine the verbs the students used for one of the target nouns, "law." Table 10 shows the verbs of the verb-law collocations that the students searched for and wrote in the essay.

Table 10. The verbs the students searched and used for verb- "law" collocations in the essay

Experimental group (Corpus users; *n* = 28)		**Control group (Dictionary users; *n* = 27)**	
verb	number of output	verb	number of output
enforce	2	have	1
pass	2	make	1
make	1		
be	1		
become	1		

Similar to the output in the post-test, apart from "make," the verbs that corpus users output in the essay are included in the top 10 verbs that they searched for verb- "law" collocations, and thus corpus users tended to use frequently searched words in the essay. In addition, corpus users tended to output frequent verb- "law" collocations, as four out of five verbs ("enforce," "make," "be," "become") were frequent verbs that appear more than 100 times in the BNC (see Figure 1). Corpus search promoted more output, especially the output of high-frequency verb- "law" collocations, than dictionary search, as well as the output in the post-test. On the other hand, for more than 60 percent of the

collocations in total, neither the corpus information nor the dictionary information was used, suggesting that the majority of the collocations the students wrote were collocations they already knew. The students used less information for the output in the essay than for answering the fill-in-the-blank questions and the output in the post-test. Because using collocations in the essays requires more spontaneous output than answering the fill-in-the-blank questions and writing collocations in the post-test, one can say that the effect of searching for collocations was stronger in the simple tasks that required less spontaneous output and weaker in the creative tasks that required more spontaneous output.

Discussion

The author showed the effects of searching and seeing corpus data on language learning by focusing on how the students used the data they accessed. Although corpus data can be used for more difficult and creative learning tasks, it is realistic for Japanese EFL learners to simply search and see the data in classroom settings. The author also argued that this study is not a simple comparison of corpus use and dictionary use. The effects of different reference materials, as well as those of different tasks, were considered to investigate how corpus data contributed to language learning.

Let us return to the research questions. The first was, "Do corpora and dictionaries have different effects on memorizing collocations?" The author's answer is negative. While the control group outperformed the experimental group on the post-test, corpus use and dictionary use did not have significantly different effects on memorizing collocations. The second research question was, "Do corpora and dictionaries have different effects on outputting of collocations?" The author's answer is positive. Corpus use promoted more output than dictionary use. Corpus use contributed to both writing collocations on the post-test and using collocations creatively in essays more than dictionary use. The third research question was, "Do corpus users and dictionary users access and process different information?" The author's answer is positive. Corpus users accessed fewer but more frequent collocations and used more of the information they accessed than dictionary users.

The results suggest that the difference in accessing and processing the information on the target collocations had different effects on their output. Corpus users' search for fewer collocations than dictionary users suggests that corpus users took more time to process the information on each collocation than the dictionary users. This would promote corpus users' deeper understanding of the target collocations, considering that deep processing better promotes recognition and recall of learned foreign words than does shallow processing (Bird, 2012). Moreover, corpus data contain more context information on target collocations than dictionary data. Exposure to rich context information would contribute to the accurate inductive inference of the target collocations' usage, and this would also promote learners' deeper understanding of the target collocations. The results also suggest that frequency information in the corpus

data helped the students focus on high-frequency collocations, considering that corpus users searched fewer but more frequent collocations than dictionary users. Since high-frequency collocations are often used in the real world and thus L2 learners should give priority to learning them, one can say that corpus search promotes L2 learning for practical use more than dictionary search. In addition, learners' focus on high-frequency collocations would help them output natural collocations because they are the collocations that native speakers often use. Therefore, it can be concluded that a deeper understanding of the target collocations' usage through the corpus's context and frequency information promoted the output of the target collocations.

However, both corpus and dictionary users output very few collocations using the target nouns – on average, less than one collocation per essay. Therefore, although the BNC users used more collocations than the dictionary users, one cannot safely state that using the BNC was effective enough to promote students' use of collocations in their essays. Since the number of collocations output in the students' essays was lower than that in the post-test, we may say that it was more difficult for students to output collocations in their essays than in the tests. That is, the creative use of collocations was more difficult than writing collocations alone.

As for memorizing collocations using the target words, the results suggest that the difference in accessing and processing the information on the target collocations did not have different effects on memorization. That is, for simple memorization of the target collocations, the students did not necessarily need the context and frequency information that the corpus provides, which had helped them output more collocations than dictionary information. Considering that dictionary users looked up more collocations than corpus users, one may say that dictionary use is time-saving and thus it could be a good option for memorizing them in a limited time.

Since it is suggested that corpus use has different effects on different tasks and has its strength in learning L2 collocations, corpus use should focus on the tasks for which corpus use is especially helpful. Considering that corpus use especially contributed to learners' output of the target collocations, corpus use in L2 writing classrooms would be a practical option. Thus, DDL would be a practical option in L2 classrooms, and teachers' understanding of the effects of corpus use on L2 learning is needed.

This study highlights one aspect of the effects of corpus use on L2 vocabulary learning, while it is not clear whether the results can be generalized due to the small number of participants. More research in this area is needed to prove the effects of corpus use on learning L2 collocations.

Conclusion

This study raised the issue of proving the effects of DDL in L2 classrooms. The author especially examined the effects of corpus use on L2 collocation learning in terms of fill-in-the-blank and output tasks. After instruction in how to use the BNC corpus, the students performed these tasks, and the pre-test and post-test

were conducted to see whether corpus use was effective in learning L2 collocations. The effects of corpus use were compared to the effects of dictionary use by the control group. Corpus users output significantly more collocations in the post-test and essays than dictionary users, as a deeper understanding of the target collocations' usage through the corpus's context and frequency information promoted the output of the target collocations. However, there was no significant difference in the average scores of the post-test fill-in-the-blank questions between corpus users and dictionary users, although the control group outperformed the experimental group. Since corpus use has different effects on different tasks, its use should focus on the tasks for which it is especially effective to maximize its strength. Thus, effective DDL in L2 classrooms requires teachers to understand the effects of corpus use and provide learners with instructions on how to use the corpus wisely.

Although this research suggests some positive effects of corpus use on L2 learning, many problems remain unclear. For effective and more varied corpus use to promote L2 learning, more research is needed to investigate the different effects that corpus use would have on tasks that were not examined in this study and how the corpus can best be used. In addition, we cannot say whether learners' proficiency levels influence task performance because this study did not have learners at different proficiency levels. Since some learners have difficulty in searching and interpreting the corpus information and thus proficiency levels are important learner variables in DDL, the effects of different proficiency levels on L2 learning with corpus use could be considered in the research design. Although further research is needed to prove the various effects of corpus use, the author hopes that the results of this study will provide insight into DDL for L2 learning to promote DDL as a practical option in L2 language classrooms.

References

Bird, S. (2012). Expert knowledge, distinctiveness, and levels of processing in language learning. *Applied Psycholinguistics*, *33*, 665–689. https://doi.org/10.1017/S014271641100052X

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). Frankfurt: Peter Lang.

Boulton, A. (2009a). Corpus for all? Learning styles and data-driven learning. In M. Mahlberg, V. González-Díaz, & C. Smith (Eds.), *5th Corpus Linguistics Conference*. Liverpool: UCREL.

Boulton, A. (2009b). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, *35*(1), 81–106.

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

Boulton, A. (2015). Applying data-driven learning to the web. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 267–295). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.69.13bou

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

Chujo, K., Kobayashi, Y., Mizumoto, A., & Oghigian, K. (2016). Exploring the effectiveness of combined web-based corpus tools for beginner EFL DDL. *Linguistics and Literature Studies*, *4*(4), 262–274. https://doi.org/10.13189/lls.2016.040404

Chujo, K., Oghigian, K., & Akasegawa, S. (2015). A corpus and grammatical browsing system for remedial EFL learners. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 109–128). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.69.06chu

Flowerdew, L. (2010). Using corpora for writing instruction. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 444–457). London: Routledge.

Flowerdew, L. (2015). Data-driven learning and language learning theory: Whither the twain shall meet. In A. Leńko-Szymańska, and A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 15–36). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.69.02flo

Hadley, G., & Charles, M. (2017). Enhancing extensive reading with data-driven learning. *Language Learning & Technology*, *21*(3), 131–152. https://doi.org/10125/44624

Hirata, E. (2019). The development of a multimodal corpus tool for young EFL learners: A case study on the integration of DDL in teacher education. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpus and DDL for pre-tertiary learners* (pp. 88–105). London: Routledge.

Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. *ELR Journal*, *4*, 1–16.

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

Leech, G. (1997). Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 1–23). London: Routledge.

Leńko-Szymańska, A. (2014). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL*, *26*(2), 260–278. https://doi.org/10.1017/S095834401400010X

Leńko-Szymańska, A. (2017). Training teachers in data-driven learning: Tackling the challenge. *Language Learning & Technology*, *21*(3), 217–241. http://hdl.handle.net/10125/44628

Leńko-Szymańska, A., & Boulton, A. (2015). Introduction: Data-driven learning in language pedagogy. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 1–14). Amsterdam: John Benjamins.

Mizumoto, A., & Chujo, K. (2015). A meta-analysis of data-driven learning approach in the Japanese EFL classroom. *English Corpus Studies*, *22*, 1–18.

Mizumoto, A., Chujo, K., & Yokota, K. (2016). Development of a scale to measure learners' perceived preferences and benefits of data-driven learning. *ReCALL*, *28*(2), 227–246. https://doi:10.1017/S0958344015000208

Mueller, C., & Jacobsen, N. (2016). 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:10.1017/S0958344015000142

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.

Piotrowski, T. (1989). Monolingual and bilingual dictionaries: Fundamental differences. In M. Tickoo (Ed.), *Learners' dictionaries: State of the art* (pp. 72–83). Singapore: SEAMEO Regional Language Centre.

Saeedakhtar, A., Bagerin, M., & Abdi, R. (2020). The effect of hands-on and hands-off data-driven learning on low-intermediate learners' verb-preposition collocations. *System*, 91, 102268. https://doi.org/10.1016/j.system.2020.102268

Satake, Y. (2020a). How error types affect the accuracy of L2 error correction with corpus use. *Journal of Second Language Writing*, *50*, 100757. https://doi.org/10.1016/j.jslw.2020.100757

Satake, Y. (2020b). The effects of corpus consultation on learning English collocations. *Journal of Corpus-based Lexicology Studies*, *2*, 13–30. http://www.lib.kobe-u.ac.jp/infolib/meta_pub/G0000003kernel_81011989

Sinclair, J. (1997). Corpus evidence in language description. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 27–39). London: Routledge.

Smith, L. C., & Mare, N. N. (2011). *Topics for today* (4th ed.). Boston: Heinle.

Tono, Y. (2015). Kopasu no eigo kyoiku eno oyo (Application of corpora to English language teaching). In Tono, Y. (Ed.), *Kopasu to eigo kyoiku* (Corpus and English education) (pp. 1–16). Tokyo: Hitsuji Shobo.

Tribble, C. (2015). Teaching and language corpora: Perspectives from a personal journey. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 37–62). Amsterdam: John Benjamins. https://doi:org/10.1075/scl.69.03tri

Vyatkina, N. (2020). Corpus-informed pedagogy in a language course: Design, implementation, and evaluation. In M. Kruk & M. Peterson (Eds.), *New technological applications for foreign and second language learning and teaching* (pp. 306–335). Hershey, PA: IGI Global. https://doi:org/10.4018/978-1-7998-2591-3.ch015

Weblio English-Japanese Dictionary and Japanese-English Dictionary. (2018). https://ejje.weblio.jp/

Widdowson, H. (1990). *Aspects of language teaching*. Oxford, UK: Oxford University Press.
Wilson, J., Hartley, A., Sharoff, S., & Stephenson, P. (2010). Advanced corpus solutions for humanities researchers. In R. Otoguro, K. Ishikawa, H. Umemoto, K. Yoshimoto, & Y. Harada (Eds.), *Proceedings of PACLIC 24* (pp. 769–778). Sendai: Tohoku University.