The Effects of Corpus Consultation on Learning English Collocations

Satake, Yoshiho

The Effects of Corpus Consultation on Learning English Collocations

SATAKE Yoshiho (Surugadai University)

satake.yoshiho@surugadai.ac.jp

英語コロケーション学習に及ぼすコーパス参照効果

佐竹　由帆（駿河台大学）

Abstract

The aim of this study was to investigate the effects of corpus consultation on learning L2 collocations. The two-group (experimental vs. control) pre-post design was used. The participants were sixty Japanese EFL learners at the CEFR B1 to B2 level. The target word was "marrow" and how students learned collocational knowledge of the word through a corpus or dictionaries was investigated. The results show that the corpus was significantly more effective in improving the learners' word associations and output of collocations than the dictionaries, while both the corpus and the dictionaries were significantly effective in memorizing collocations. The corpus users looked up fewer but more frequent collocations than the dictionary users. It is suggested that corpus consultation promoted the learners' deeper understanding of the target word usage because of a corpus's rich contextual and frequency information.

Keywords

Data-Driven Learning (DDL), Corpus Consultation, L2 Collocation Learning

1. Introduction

Recently various corpus tools have been developed and helped identify diverse linguistic features (e.g. collocations, grammatical patterns, keywords) (Leńko-Szymańska & Boulton, 2015). Corpora have affected studies in diverse areas and language teaching is one of them. There are two ways of using corpora for language pedagogy. One is direct use and students themselves directly use corpora and analyze the data to learn the target language. Johns (1991) coined the term data-driven learning (DDL) for direct use of corpus for language learning. Flowerdew (2010) states that DDL promotes learners' acquisition of phraseological patterns (i.e. collocations, colligations

and semantic preferences and prosodies), as it is not easy for learners to access them in dictionaries or grammar books. The other way of using corpora is indirect use that involves making dictionaries and language textbooks by using corpus data. Collins COBUILD English Language Dictionary (1987) is a pioneering example of indirect corpus consultation. Although indirect use of corpora has prevailed and various corpus-based dictionaries and textbooks have been published, direct corpus consultation has not prevailed (Flowerdew, 2010, Leńko-Szymańska & Boulton, 2015).

There are some reasons that DDL has not prevailed in L2 classrooms. Leńko-Szymańska and Boulton (2015) discuss that teachers tend to think DDL is difficult to introduce into classrooms, as it is not easy for learners to use and analyze corpus data. Studies have often stated that insufficient language skills make it difficult for low-proficiency learners to use corpus data (e.g., Chujo, Oghigian, & Akasegawa, 2015). Besides, teachers have made a complaint that they do not have enough opportunities to get instruction on how to use corpus information effectively (Heather & Helt, 2012; Leńko-Szymańska, 2014). To promote DDL in language classrooms, more empirical research is needed to prove the effects of DDL and persuade language teachers to use DDL.

The author's previous study (Satake, 2014) compares the effects of corpus use and dictionary use and finds out that corpus use promotes more output of collocations than dictionary use, while dictionary use promotes more search and memorization of collocations than corpus use. However, the students' L1 could have favorably affected the results of the dictionary users because they mainly used English-Japanese dictionaries and the author gave them the Japanese-English translation test to test their knowledge of collocations. To prove the specific effects of corpus consultation, further research is needed.

Thus this research addresses the issue of how and where corpus consultation could contribute to L2 vocabulary learning. The author deals with how DDL can contribute to improving word associations as well as memorizing and outputting L2 collocations, which the author examined in the above-mentioned study. Unlike the previous study, the author did not use the Japanese-English translation test in this study so that the learners' L1 would not favorably influence the results of the dictionary users, who chiefly used English-Japanese dictionaries.

The author starts with a review of the research on the effects of DDL in L2 classrooms. Then the author describes the research questions and the research design. The author finally shows the results and discusses the implications of this research.

2. Literature Review

As communicative language teaching (CLT) has promoted learner-centered learning, information and communication technology (ICT) has made the role of the teacher more learning-focused than teaching-focused (Boulton 2009). Computer-based DDL utilizes ICT and learners are exposed to a vast amount of authentic data from corpora and encouraged to investigate example sentences and detect the usage pattern (Boulton, 2009). According to Johns (1991), learners go through three stages in DDL—identify, classify, and generalize. Considering that they are convertible to Carter and McCarthy's (1995) illustration, interaction, and induction (Ishikawa, 2008), they could take over the traditional approach of presentation, practice, and production and contribute to more learner-centered teaching (Hayama 2015).

However, not many teachers have adopted a DDL approach in spite of its strength (e.g., Flowerdew, 2010). Teachers are sometimes reluctant to use corpora in classrooms because they are not user-friendly enough (Kosem, 2008) and thus it is not easy for learners to use them (e.g., Chujo, Utiyama, & Nishigaki, 2007). Lack of knowledge about corpus consultation also prevents teachers from using corpora and thus teachers need corpus instruction (Mukherjee, 2004). Learners, as well as teachers, need corpus instruction. In addition, teachers are not convinced to use corpora in classrooms because many things remain unclear about the effects of corpus consultation.

Previous studies have proved some effects of DDL. Mueller and Jacobsen (2015) compared the effects of L2 error correction with corpus use and those of dictionary use and found out that corpus use promoted understanding prepositions and light verbs more than dictionary use, while corpus use and dictionary use do not show different effects on understanding register. Luo and Liao (2015) also compared the effects of corpus and dictionary use on L2 error correction and discovered that corpus consultation promoted more accurate correction and fewer errors in new essays than dictionary use.

The author's previous group research also examined the effects of corpus consultation. Tono, Satake, and Miura (2014) explored the effects of corpus consultation on L2 error correction. 93 Japanese EFL learners did a timed essay task without consulting reference materials and were given feedback for two errors by the authors. After three weeks, the learners performed the revision task consulting the British National Corpus (BNC) or dictionaries of their choice and then the authors grouped the 188 errors into the three categories (omission, addition, and misformation). They discovered that corpus consultation contributed to correcting omission and addition errors accurately, while it did not contribute to correcting misformation errors enough.

The author also explored the effects of corpus consultation on other areas than error

correction. As mentioned above, Satake (2014) discovered that DDL contributed to outputting collocations, although dictionaries contributed to searching and memorizing collocations. Satake (2015) compared the effects of corpus and dictionary use on different tasks (i.e., gap-filling, outputting phrases, semantic prosody and analogy tasks). She found out that corpus consultation promoted more outputs than dictionary use and that the corpus users, especially the learners with more English knowledge, were better at semantic prosody and analogy tasks than the dictionary users.

Although the above-mentioned research shows some positive effects of DDL on L2 learning such as the tasks of error correction, vocabulary learning, and output, there are still unclear problems. While Flowerdew (2010) says that corpus consultation promotes students' acquiring phraseological patterns such as collocations more than dictionaries because it is not easy to access them in dictionaries, how and why corpus consultation contributes to learning collocation is not clear enough. Since "language knowledge is collocational knowledge" (Nation, 2001: 318) and thus language acquisition requires learners to understand collocations, we can say that more studies are needed to prove the effects of corpus consultation on learning L2 collocations.

3. Methods

3.1 Aim and Research Questions

The aim of this research was to examine the effects of corpus consultation on learning L2 collocations. This research hypothesizes that accessing and analyzing many usages of the target word(s) in a corpus contribute to memorizing and outputting collocations of the target word(s) and improve word associations, as seeing many example sentences promotes inductive inference of the appropriate pattern. The author compared the effects of corpus consultation with the effects of dictionary use so that she could examine how corpus consultation affected learning L2 collocations. The research questions were as follows:

(1) Do a corpus and dictionaries produce different effects on memorizing collocations?

(2) Do a corpus and dictionaries produce different effects on learners' word associations?

(3) Do a corpus and dictionaries produce different effects on the output of collocations?

(4) Do corpus and dictionary users access and process different information?

3.2 Participants

The participants were two classes of intermediate Japanese learners of English (60) at a private university in Tokyo, some of whom had passed the EIKEN Test of Practical

English Proficiency Grade 2, which is roughly equivalent to B1 in the Common European Framework of Reference for Languages (CEFR). A few students reached the CEFR B2 level. The participants were taking a compulsory English reading course taught by the author. Before the author began the study, she told the aim of this research in Japanese. The author also told that the participants' anonymity was assured, that their participation was not compulsory, and that their participation or nonparticipation would not influence their grades. The author asked the participants for permission to use their data and received written permission from 60 students.

3.3 Instruments

3.3.1 The Corpus

The Corpus of Contemporary American English (COCA) was used for this study because it is a large, balanced corpus, which can help English language learners learn authentic English. The COCA is an approximately 600-million-word corpus of American written and spoken English since 1990. Before the tasks, the participants who were assigned to the COCA were provided with a 20-minute instruction on how to use the corpus tools, how to search for the collocations, and how to sort and interpret the concordance lines.

3.3.2 Dictionaries

The participants could use any dictionaries of their choice. The majority of them used English-Japanese dictionaries and approximately half of them used the online Weblio English-Japanese Dictionary that has more than ten million words (2018). Almost all of them chose online dictionaries.

3.3.3 Task

A ten-minute timed task was used in this research. Before the first task, special instructions were provided to demonstrate how to use the corpus tools and how to search and analyze the corpus data. The target collocations were the collocations using "marrow." The words "marrow" was chosen as the target word because it was often used in the article the participants were reading in class. Another reason for choosing "marrow" was that the author thought "marrow" was appropriate for the research to test the effects of corpus consultation because it is not a high-frequency word and thus the majority of the participants would not have enough collocational knowledge of it, which could interfere with the research results. The article used for this research was about the ethics of planned pregnancy to conceive a child whose marrow would be a close match for her

sister, who suffered from leukemia (Smith & Mare, 2011). For the task, after searching the article for the collocations using "marrow", the participants were told to look up three or more collocations for the target word, for which the experimental group with 29 students used the COCA to search for the target collocations, see the frequency list of co-occurrence words, and read the KWIC concordance lines. The control group with 31 students used dictionaries. Next, all participants typed the collocations they had searched.

3.3.4 Pre- and Post-Tests

Five-minute pre and post-tests were used in this research. Before the task, the participants did a pre-test on the target noun, which had five multiple-choice fill-in-the-blank questions and a word-association quiz. As for fill-in-the-blank questions, there was a blank before or after the target noun "marrow", and the participants were told to fill in each blank with an appropriate word, an example of which is shown below.

(2) marrow (   )
1 transfer
2 transmit
3 transplant
4 transpose

Among the five collocations in the questions, two were used in the textbook, and the other three collocations were high-frequency collocations in the COCA. High-frequency collocations were selected for the questions so that the participants could easily find the collocations in the COCA or in the dictionaries when they searched them later. One point was given for a correct answer and thus full marks were five. As for a word-association quiz, the participants were asked to write words, with which they associate "marrow". Word associations did not necessarily have to be collocations. For instance, "leukemia" was regarded as a relevant word association. The author gave one point to each word association, which she considered as relevant. Immediately after the task, a post-test was conducted, which was almost the same as the pre-test except that the students were required to write as many phrases as possible using the target noun, "marrow". A collocation was regarded as appropriate if it appeared in the COCA.

3.4 Procedure

The following procedure was adopted:

1. The pre-test was conducted (five minutes).

2. The instruction was provided for COCA (twenty minutes).

3. A task session was held. The timed task for finding and typing collocations using "marrow" was given (ten minutes).

4. Immediately after the task session, the post-test was conducted (five minutes).

A two-group (corpus users vs. dictionary users) pre-post design was used to examine the effects of corpus consultation for learning collocations. Both groups were provided with five minutes to take the pre-test. Before doing the task, the experimental group was provided with twenty minutes' instruction on how to use the corpus tools and analyze the corpus data. Then the participants were provided ten minutes to do the task. The experimental group used the COCA to do the task, and the control group used dictionaries they chose. Immediately after the task, both groups took the post-test.

The results from the pre-test and the post-test were then compared to investigate the different effects on the retention and output of the target collocations between the two groups. The number of the collocations each group searched was also compared, categorized by frequency of collocations. Five collocated words that appear within three words before and after "marrow" more than fifty times in the COCA (i.e., "bone", "transplant", "cell", "donor", "blood") were regarded as high frequent.

4. Results

4.1 Effects of Corpus Consultation

4.1.1 Fill-in-the-Blank Questions

Table 1 summarizes the average marks for the fill-in-the-blank questions in the pre and post-tests, categorized by reference material type. To judge whether there were any significant differences in the marks between the reference materials and between the pre-test and the post-test, an analysis of variance (ANOVA) was conducted for the marks of all the students. A significant difference was found both between the reference materials (F (1, 58) = 5.00, p < .05, η2 = .10) and between the tests (F(1, 58) = 200.66, p < .01, η2 = .69). There was an interaction effect between the reference materials and the tests (F(1, 58) = 5.18, p < .05, η2 = .02). Between the pre- and post-tests, the simple main effect test shows there was a significant difference in the marks by both the corpus users (F(1, 58) = 70.68, p< .01) and the dictionary users (F(1, 58) = 135.16, p < .01) and thus it is suggested that both corpus and dictionary use helped the learners retain the target collocations. Between the corpus users and the dictionary users, the simple main effect

test shows there was a significant difference in the marks of the pre-test (F(1, 58) = 8.63, p< .01), while it does not show there was a significant difference in the marks of the post-test (F(1, 58) = 1.42, p = .24). The significant difference in the marks of the pre-test between corpus and dictionary users suggests that dictionary users' English knowledge was less than corpus users, contrary to the supposition of the author. Therefore, it is concluded that both corpus and dictionary use helped the learners retain the target collocations as well as they did in Satake (2014), while it remains unclear about whether there were the different effects between them, unlike Satake (2014) suggests that dictionary use was more effective for collocation retention than corpus use. Since in Satake (2014) Japanese-English translation test could have favorably affected the results of English-Japanese dictionary users, this study suggests that there was no different effect between corpus use and bilingual dictionary use when the learners' L1 was not used for the test.

Table 1. *Average marks for the fill-in-the-blank questions*

	Experimental group (Corpus users; n = 29)	Control group (Dictionary users; n = 31)
Pre-test	3.4	2.9
Post-test	4.2	4

4.1.2 Word associations

Table 2 summarizes the average number of words with which they associated "marrow" in the pre- and post-tests, categorized by reference material type. To judge whether there were any significant differences in the number for the word association quiz between the reference materials and between the pre-test and the post-test, ANOVA was conducted for the number for the quiz of all the students. A significant difference was found both between the reference materials (F(1, 58) = 63.31, p < .01, η2 = .79) and between the tests (F(1, 58) =18.64, p < .01, η2 = .05). There was also an interaction effect between the reference materials and the tests (F(1, 58) = 9.57, p < .01, η2 = .02). Between the corpus users and the dictionary users, the simple main effect test shows there was a significant difference both in the marks of the pre-test (F(1, 58) = 36.49, p< .01) and in those of the post-test (F(1, 58) = 72.74, p< .01). The significant difference in the marks of the pre-test between corpus and dictionary users again suggests that dictionary users' English knowledge was less than corpus users (see 4.1.1) and that this might have had some influence on this research. Between the pre- and post-tests, the simple main effect test shows there was a significant difference in the marks by the corpus users (F(1, 58)

= 27.47, p< .01), while it does not show there was a significant difference in those by the dictionary users (F(1, 58) = 0.75, p = .39). Therefore, it was concluded that only corpus consultation promoted word association of the target word.

Table 2. *Average number for the word association quiz*

	Experimental group (Corpus users; n = 29)	Control group (Dictionary users; n = 31)
Pre-test	4.4	3.1
Post-test	5	3.2

Then let us qualitatively analyze the words the participants used for word association quiz. Table 3 shows the top five words used for the word association quiz. For the pre-test, both the corpus users and the dictionary users mainly used the words that appear in the article they were reading. However, for the post-test, while the dictionary users mainly used the words in the article again and there was only one word which is not included in the article (i.e., "cancer"), the corpus users used different words, which are frequently used with the target word "marrow" in the COCA corpus (see 3.4). It is suggested that the information the corpus users accessed through the collocation-searching task helped them improve their word associations.

Table 3. *Top five words used for the word association quiz*

	Experimental group (Corpus users; n = 29)		Control group (Dictionary users; n = 31)	
Rank	Pre-test	Post-test	Pre-test	Post-test
1	baby	bone	baby	bone
2	save	transplant	life	baby
3	life	cell	hospital	life
4	bone	blood	pregnancy	cancer
5	red	donor	bone	hospital

4.1.3 Output from the Post-Test

Table 4 summarizes the average collocation outputs in the post-test. In the post-test, both groups were required to write as many collocations for the target noun "marrow" as possible. The number of appropriate collocations was then counted. On average, the corpus users were able to produce a greater number of correct collocations than the dictionary users. To judge whether there were any significant differences in the number

of the collocation outputs between the experimental group and the control group, the Mann-Whitney U-test was applied for the number of the outputs because the data did not show a normal distribution. Significant differences were found (z = −5.94, p < .01, r = .77) and thus indicated that corpus use significantly promoted more outputs of the target collocations in the post-test than dictionary use.

Table 4. *Average number of the target-collocation outputs in the post-test*

	Experimental group (Corpus users; n = 29)	Control group (Dictionary users; n = 31)
Total	2.4	1.3

Next let us qualitatively analyze the words the participants used for collocation output. Table 5 shows the top five words used for collocation output in the post-test. The words share the same tendency as those for word associations. While the dictionary users mainly used the words in the article again and the five words were included in the article, the corpus users used only two out of five words in the article ("bone", "transplant") and the other three words are frequently used with the target word "marrow" in the COCA corpus (see 3.4). It is suggested that the information the corpus users accessed through the collocation-searching task helped them promote their collocation output.

Table 5. *Top five words used for collocation output in the post-test*

Rank	Experimental group (Corpus users; n = 29)	Control group (Dictionary users; n = 31)
1	bone	bone
2	transplant	baby
3	donor	transplant
4	cell	match
5	blood	draw

In conclusion, the results show the same tendency as the results in Satake (2014) and it was suggested that corpus use assisted the learners output the target collocations than dictionary use. However, since the number of the output of collocations in the post-tests was lower than the average mark for the fill-in-the-blank questions in the pre-test and the post-test, it is suggested that it was more difficult for the learners to output collocations than to fill in an appropriate word into each blank. It can be concluded that using collocations creatively was more difficult than just writing collocations.

4.2 Average Number of Collocations Searched for

Table 6 summarizes the average number of collocations that the learners looked up for 10 minutes, categorized by frequency of collocations. The results show the same tendency as the results in Satake (2014), and the dictionary users searched for a greater number of collocations than the corpus users, which may have been because it was difficult for the learners to search and interpret the corpus information. To judge whether there were any significant differences between the number of collocations searched for by the corpus users and the dictionary users, the chi-square test and the residual analysis were conducted for the total number of collocations that the participants searched. A significant difference was found ($\chi^2(1) = 52.86$, $p < 0.01$, $V = .43$), confirming that the corpus users searched for more high frequent collocations and fewer low frequent collocations than the dictionary users.

Table 6. *Average number of the collocations that the learners searched for*

	Experimental group (Corpus users; n = 29)	Control group (Dictionary users; n = 31)
High frequent collocations	3	2.5
Other collocations	0.5	3.6
Average total	3.5	6.1

Next, let us qualitatively analyze the words the participants looked up during the collocation-search task. Table 7 shows the top five words that the participants searched for. The words searched by the corpus users and those searched by the dictionary users shared three high frequent co-occurrence words with "marrow" and thus one can say that both types of materials provided the learners with practical and useful information. However, there was a difference between them on how they used the information they accessed. While the corpus users used four out of five words for word associations and collocation output, the dictionary users used only one word for word associations and two words for collocation output (see Table 3 & 5). The corpus users' more use of the information they accessed suggests that exposure to a greater number of high frequent collocations promoted word associations and collocation output because concordance lines show a much greater number of high frequent co-occurrence words and collocations than example sentences in dictionaries.

Table 7. *Top five words that the learners looked up during the collocation-search task*

Rank	Experimental group (Corpus users; n = 29)	Control group (Dictionary users; n = 31)
1	bone	bone
2	transplant	transplant
3	donor	speech
4	cell	transplantation
5	transplantation	vegetable

Considering that the corpus users searched for fewer collocations in total than the dictionary users, it is also suggested that the corpus users may have spent more time on each colocation than the dictionary users, which could have contributed to the more output of collocations. Since learning high frequent collocations is more encouraged than learning low frequent collocations, this result suggests that corpus use could promote more efficient learning of high frequent collocations than dictionary use.

For pedagogical implications, to promote more output of collocations, corpus use should be encouraged since a corpus provides learners with exposure to more high frequent collocations than dictionaries. It is also suggested that teachers should provide instruction on the number of collocations learners should search so that they can take enough time to process the information they access.

4.3 How the Learners Used the Information They Accessed

To examine how the learners used the information they accessed, the author analyzed how the collocations they searched for during the task influenced the marks of fill-in-the-blank questions and the number of outputs in the post-test. The author did not analyze how the information they accessed influenced word associations because the participants did not directly search for the words with which they associate the target word "marrow".

4.3.1 Fill-in-the-Blank Questions

Table 8 summarizes how the collocation searches influenced the answers to the fill-in-the-blank questions in the post-test. The corpus users' searches promoted greater use of the information they accessed than the dictionary users'. More than 30% of the total answers by the corpus group was appropriately used and contributed to the correct answers, while less than 25% of the total answers by the dictionary group contributed to the correct answers. On the other hand, considering both the corpus users and the dictionary users gave very few wrong answers to the collocations that they searched for,

it can be said that both corpus use and dictionary use contributed to accurate retention of collocations. This result suggested that the corpus was more efficiently used to retain the collocation information than the dictionary. However, for more than 50% of the total answers, neither the corpus information nor the dictionary information was used. Although the main results show the same tendency as the results in Satake (2014), the results of this research show much fewer wrong answers to the collocations that they searched for than the results in Satake (2014). The difference of the target collocations could have influenced the results: in Satake (2014), the participants were asked to focus on the verb-target noun collocations, while in this research the participants were asked to focus on the collocations using the target word and there was no restriction of parts of speech of co-occurrence words, which could have made collocation learning easier. The target words could have also influenced the results: the target word "marrow" in this research had lower but the stronger collocations than the target words "law" and "pregnancy", and thus the participants of this research could have focused on the high-frequency collocations that appeared in the tests when they did the collocation-search task.

Table 8. *Effect of the collocation searches on the fill-in-the-blank answers*

		Experimental group (Corpus users; n = 29)	Control group (Dictionary users; n = 31)
Searched	Correct	33.1%	24.5%
	Wrong	1.4%	0.6%
Not searched	Correct	51.0%	54.8%
	Wrong	14.5%	20.0%

4.3.2 Output from the Post-Test

Table 9 summarizes the average collocation output in the post-test, categorized by whether the collocation was searched for or not when the participants did the collocation-search task. To judge whether there were any significant differences in the number of the collocation output between the collocations that were searched for and those that were not during the collocation-search task, ANOVA was conducted. A significant difference was found ($F(1, 58) = 19.28$, $p < .01$, $\eta^2 = .11$), confirming that the participants output significantly more collocations for which they searched than those for which they did not search. The corpus users' more output of collocations they searched for during the task suggests that the corpus users more efficiently used the information they accessed. In addition, the corpus users' more output of collocations they did not type

during the task suggests the possibility that there was some output of collocations that they looked over and used the information even though they did not type them during the collocation-search task. The results of this research are in contradiction with those in Satake (2014), which show that the participants output fewer collocations for which they searched than those for which they did not search. More studies are needed to discover how learners use the corpus information they accessed and the effects of corpus use for collocation output.

Table 9. *Average number of the target-collocation outputs in the post-test*

	Experimental group (Corpus users; n = 29)	Control group (Dictionary users; n = 31)
Searched	1.3	0.8
Not searched	1.1	0.5

5. Discussion

The effects of using corpus data for language learning were examined by focusing on how the learners used the data they accessed, compared with the effects of using dictionary data they accessed. This research was, however, not only a comparison of corpus use and dictionary use but also how the use of the corpus data promoted language learning for different tasks.

The first research question was about whether corpus and dictionary use had different effects on memorizing collocations. The answer to this question remains unclear, while it is suggested that both corpus use and dictionary use promoted collocation retention.

The second research question was about whether corpus and dictionary use had different effects on learning word associations. The answer to this question was that only corpus use contributed to improving word association of the target word, as only corpus use promoted a significantly greater number of word associations in the post-test than in the pre-test. However, since the corpus users' top five words used for word association quiz and the words used for collocation output in the post-test shared all of the five words, one may say that it is unclear about the difference between word association output and collocation output.

The third research question was about whether corpus and dictionary use had different effects on collocation output. It was concluded that corpus use promoted more output of collocations than dictionary use.

The fourth research question was about whether corpus users and dictionary users

accessed and processed different information. It is suggested that the corpus users more efficiently used the information for language learning than the dictionary users because the corpus users accessed fewer but more high frequent collocations, and used more of the information they accessed than the dictionary users.

Judging from the above results, how the learners accessed and processed the information of the target collocations influenced how they output the target collocations. It is suggested that the frequency information in the corpus assisted the participants to pay attention to and then output high-frequency collocations because the corpus users looked up fewer but more high-frequency collocations than the dictionary users. It can be said that corpus frequency information promoted more practical L2 learning, considering that L2 learners should give priority to learning high-frequency collocations, which are used more than low-frequency collocations.

Besides, the result that the corpus users looked up fewer collocations than the dictionary users indicates that the corpus users spent more time to process the information about each collocation than the dictionary users. This could contribute to a more thorough understanding of the target collocations. Moreover, since concordance lines of the target word show richer contextual information than dictionaries, this exposure to more context information could improve learners' understanding of the target collocations and thus lead to a correct inductive inference about the usages of the target collocations. It is also suggested that a deeper understanding through the exposure to more context information of the target word improved word associations.

On the other hand, it remains unclear whether corpus use and dictionary use had different effects on memorizing the target collocations, while it is concluded that both helped the learners memorize them. In addition, we are not sure about how different parts of speech of co-occurrence words and different learners' L2 proficiency have different effects on DDL, how strong and week collocations could have different effects on learning them, and what relationship between word associations and collocations is. Besides, some results of this research—how the collocation information the learners searched promoted accuracy in fill-in-the-blank questions and collocation output— were different from those of Satake (2014). To prove the various effects that corpus use would have on L2 vocabulary learning, more studies in this field are needed.

To sum up, the strength of using corpora for learning collocations lies in its effectiveness in promoting learners' word associations and output of collocations. Corpora should be used for the tasks for which it is particularly helpful. For effective DDL in L2 classrooms, teachers' understanding and preparation of appropriate tasks for corpus use are needed.

6. Conclusion

This research explored the effects of corpus use on learning English collocations. The two-group (experimental vs. control) pre-post design was used to assess whether corpus use was effective in learning English collocations, compared to the effects of dictionary use. In conclusion, corpus use promoted word associations and collocation output of the target word more than dictionary use because the corpus users reached more thorough understanding of the usages of the target word owing to the information of its frequency and context, while it is not clear about the different effects that corpus and dictionary use could have on collocation retention. Since it is suggested that corpus consultation had different learning effects for different tasks, effective corpus use in L2 classrooms requires teachers to understand and prepare appropriate tasks for corpus consultation.

This research suggested some positive effects of corpus consultation for L2 learning tasks, while there are some limitations. We are not sure from this research that whether corpus consultation has long-term learning effects on word associations and collocation output because this research did only the immediate post-test. To prove the long-term effects, a delayed post-test is needed in future research designs. Although further research is required to test various effects of corpus use could have on L2 learning, one can say that the results of this research highlight some strengths of DDL for L2 learning.

References

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

Carter, R., & McCarthy, M. (1995). Grammar and the spoken language. *Applied Linguistics, 16*, 141-158.

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.

Chujo, K., Utiyama, M., & Nishigaki, C. (2007). Towards building a usable corpus collection for the ELT classroom. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language* (pp. 47–69). Amsterdam: Rodopi.

Collins COBUILD English Language Dictionary. (1987). London: Collins ELT.

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

Hayama, M. (2015). Kopasu to eigo kyoju (Corpora and teaching English). In Y. Tono (Ed.), *Kopasu to eigo kyoiku* (Corpora and English education) (pp. 43-71). Tokyo: Hituzi Syobo.

Heather, J., & Helt, M. (2012). Evaluating corpus literacy training for pre-service language teachers: Six case studies. *Journal of Technology and Teacher Education, 20*(4), 415-440.

Ishikawa, S. (2008). *Eigo kopasu to gengo kyoiku: Deta toshiteno tekusuto* (English corpora and language teaching: Texts as data). Tokyo: Taishukan.

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

Kosem, I. (2008). User-friendly corpus tools for language teaching and learning. In A. Frankenberg-Garcia, T. Rkibi, M. R. Cruz, R. Carvalho, C. Direito, & D. Santos-Rosa (Eds.), *Proceedings of the 8th Teaching and Language Corpora Conference* (pp. 183–192). Lisbon: ISLA.

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.

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.

Luo, Q. & Liao, Y. (2015). Using corpora for error correction in EFL learners' writing. *Journal of Language Teaching and Research 6*(6), 1333-1342.

Mueller, C. M. & Jacobsen, N. D. A. (2015).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.

Mukherjee, J. (2004). Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In U. Connor & T. Upton (Eds.), *Applied corpus linguistics: A multi-dimensional perspective* (pp. 239–250). Amsterdam, Rodopi.

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

Satake, Y. (2014). Corpora vs. dictionaries: Their effects on learning English collocations in L2 writing tasks. Presented at the Second Asia Pacific Corpus Linguistics Conference (APCLC). Hong Kong, March 7-9.

Satake, Y. (2015). Comparison of dictionary use and corpus use: Different effects on

learning L2 phrases. In L. Li, J. Mckeown, & L. Liu (Ed.), *Proceedings of ASIALEX 2015* (pp. 222-229). Hong Kong: Hong Kong Polytechnic University.

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

Tono, Y., Satake, Y., & Miura, A. (2014). The effects of using corpora on revision tasks in L2 writing with coded error feedback. *ReCALL, 26*(2), 147-162.