



The Effect of Online English Corpora on Chinese EFL Learners' Acquisition of Vocabulary Knowledge

Tongqing Gu*

School of Foreign Languages, Chengdu University of Information Technology, Chengdu, 610225, China

*915131080@qq.com

Abstract. Large online English corpora are important resources to enhance English teaching and learning owing to their large scale, authentic data, fast speed and free online use. This study examined the role of online English corpus COCA and BNC in promoting English learners' in-depth acquisition of vocabulary knowledge, especially productive vocabulary knowledge. After a three-month teaching intervention of computer-assisted vocabulary learning, the results of the vocabulary knowledge tests of the subjects were analyzed. The findings confirmed two hypotheses: i. Computer-assisted vocabulary learning can expand the depth of L2 learners' vocabulary knowledge and promote their vocabulary competence. ii. There are differences in the depth of L2 vocabulary acquisition among learners with different L2 proficiency and vocabulary levels.

Keywords: Online English corpora, Depth of vocabulary knowledge, English majors

1 Introduction

Computer-assisted language learning (CALL) is an important research topic in the field of applied linguistics, especially in foreign language teaching and learning [1]. Large online English corpora are computerized resources to enhance English learning owing to their large scale, authentic data, fast speed and free online use [2]. Under the guidance of the theory of second language vocabulary acquisition and the theoretical framework of lexical knowledge[3], this study investigated the effect of online computerized corpora, the Corpus of Contemporary American English (COCA) and the British National Corpus (BNC), as well as the online English learning dictionaries, the Oxford English Collocation Dictionary (OED) and the Macmillan English Dictionary (MED), in promoting the acquisition of lexical knowledge of Chinese EFL (English as a foreign language) learners, so as to explore ways to promote the development of vocabulary competence, especially productive vocabulary competence, of foreign language learners.

2 Research Design

2.1 Research Hypotheses

This study aims to test two hypotheses: H1: Corpus-assisted vocabulary learning can expand L2 learners' depth of vocabulary knowledge, promote their vocabulary development. H2: Given the interconnection of vocabulary knowledge with L2 proficiency, there are differences in the acquisition of L2 vocabulary knowledge among learners with different L2 proficiency and vocabulary levels [4].

2.2 Research Participants

Participants of the study were 58 second-year English majors randomly selected from the course "Reading in English Current Text" taught by the author. Among them, there are 16 male students and 42 female students, and their age range is between 18 and 20. They have mastered basic English skills through specialized study. According to the *English Syllabus for English Majors in Colleges and Universities*, second-year English majors should recognize 5,500-6,500 words, and use 3,000-4,000 of them correctly and proficiently. Therefore, the experimental subjects were able to fulfill all the learning and testing tasks required by this study.

2.3 Research Instruments and Materials

This study employed two vocabulary tests (<http://www.lextutor.ca/tests/>). The Vocabulary Level Test (VLT) designed by Schmitt (2001) was used to measure subjects' vocabulary level. The general vocabulary is divided into four word frequency levels: 2,000, 3,000, 5,000 and 10,000 words. Each lexical frequency level contains 10 groups of words, including 5 groups of nouns, 3 groups of verbs and 2 groups of adjectives. Within each group, there are 6 words, which are required to be paired with 3 interpretive meanings. There are 30 target words in each word frequency level. Scoring is based on a scale of 1 point for a correct choice and 0 point for a wrong choice, with a total score of 120.

The Word Associates Test (WAT), designed by Read (1998), was used to test two areas of vocabulary knowledge: meaning and collocation. WAT consist of 40 target words, which are all adjectives having multiple meanings and uses. Each item has one stimulus and eight options, divided into two groups of four words each. There are one to three words in the left box that are synonymously related to the target word and one to three words in the right box that can co-occur with the target word. Subjects were asked to select 4 words from these two groups that were semantically and pragmatically related to the target word. One point was awarded for selecting one correct answer out of a total of 160 points. Reliability of the instrument reaches 0.93^[5].

The textbook used in this study is *Current News Articles for English Reading*. The texts are authentic and up-to-date, with various topics covering politics, economy, culture, education, science and technology. There are eight units in the book, each consisting of three articles on the same topic. The vocabulary exercises involve syn-

onymy, antonymy, polysemy and hyponymy of the target word to reinforce the vocabulary knowledge learned in the text [6].

2.4 Research Procedures

Prior to the experiment, VLT and WAT were administered in the experimental class. Students were then divided into high, mid, and low proficiency levels based on the scores of VLT and their General Point Average, as shown in Table 1.

Table 1. Description of proficiency levels.

Proficiency level	High GPA \geq 4, VLT \geq 5000	Intermediate GPA \geq 3.5, 5000 $>$ VLT $>$ 3000	Low GPA \leq 3, VLT \leq 3000
N	15	34	9

Experimental subjects were first trained on the use of the online corpora COCA and BNC. They then carried out various forms of data-driven learning to expand their in-depth vocabulary knowledge by utilizing the multi-level and multi-dimensional search functions provided by the online corpora: a. Use chart to learn the frequency of vocabulary usage; b. Use the balanced corpus to analyze the stylistic features; c. Master the collocation patterns by using the multi-level combination query function; d. Use the lexical comparison function to identify synonyms and near-synonyms; e. Observe and analyze concordance lines to understand the syntactic forms and grammatical functions; f. Grasp the contextual information of vocabulary through contextual co-occurrences and extended contexts; g. Use the dynamic corpus to understand the change and development of English vocabulary.

After class, students were required to search core vocabulary of the text in the online corpora COCA and BNC to observe the concordance lines and expanded context to learn about the style, contexts and other lexical information about the core vocabulary of each unit. Meanwhile, according to the topic of each unit, students were required to write a review on the current text they have read. They should try to use the core vocabulary learned when writing the review, and their work was graded and included in their regular grades.

Immediate post-test of WAT was administered soon after the experiment, and delayed post-test was conducted two weeks after the experiment to examine their acquisition of in-depth vocabulary knowledge.

3 Results and Discussion

To test the first hypothesis, paired samples T-test was conducted to compare pretest and post-test scores of WAT and the results are shown in Table 2 and Table 3.

As shown in Table 2, there is a significant difference between the pretest and post-test scores of WAT ($p = .000 < 0.05$), indicating that utilizing online corpora to assist vocabulary learning can enhance learners' depth of vocabulary knowledge. The online corpora can provide students with opportunities to acquire vocabulary

knowledge in a richer and more intuitive way. Corder (1981) has pointed out that we cannot actually 'teach' a foreign language to our students; we can only create a suitable language learning environment for them [7], and computer-assisted vocabulary learning is precisely the use of online resources to optimize the learning environment, and ultimately improves the effectiveness of vocabulary learning.

To explore further the effect of online corpora on participants' acquisition of vocabulary knowledge, paired samples T-test was conducted to compare pretest and delayed post-test of WAT and the results are shown in Table 3.

Table 2. Paired samples test

		Paired Differences				
		95%Confidence Interval of the Difference				
		Lower	Upper	t	df	Sig. (2 tailed)
Pair 1	Pretest – post-test	-13.11302	-4.99043	-4.463	57	.000

Table 3. Paired samples test

		Paired Differences				
		95%Confidence Interval of the Difference				
		Lower	Upper	t	df	Sig. (2 tailed)
Pair 1	Pretest – Delayed post-test	-6.90176	.79831	-1.587	57	.118

Table 3 shows that there was no significant difference between the pretest and the delayed post-test ($p > 0.05$). This indicates that the learners' vocabulary knowledge of the target words has not been fully transformed into long term memory. The reason for this is that the retention of vocabulary knowledge requires learners to "retrieve" and "rehearse" the target words they have learned several times. Webb (2007) holds that the number of encounters with each word is important [8]. There are many ways to retrieve and rehearse vocabulary. Teachers provide oral and written practice for students to retrieve target vocabulary in use, while students are expected to develop and implement specific vocabulary learning plans, such as using a word book to review and reinforce vocabulary learned.

To test hypothesis 2, One-Way ANOVA was conducted to compare WAT scores of participants with different L2 proficiency and vocabulary levels, and the results are shown in Table 4 and Table 5.

Table 4 shows the comparison of the participants' scores in the post-test of vocabulary depth knowledge between the three proficiency groups. The results show that the difference between the three groups is significant ($p = .000 < 0.05$).

One-way ANOVA of the overall scores of the three proficiency groups in the delayed post-test of vocabulary depth knowledge in Table 5 also proved that the difference between the three groups is significant ($p = .003 < 0.05$).

To summarize, the results confirmed that there are significant differences in the depth of second language vocabulary knowledge acquisition among learners with different proficiency levels and vocabulary sizes. Liu (2001) also found through an empirical study on the acquisition of word meanings and affixes by learners of different

L2 levels that L2 vocabulary knowledge, L2 proficiency level and vocabulary size are interrelated [9].

Table 4. ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4661.543	2	2330.772	22.720	.000
Within Groups	5642.181	55	102.585		
Total	10303.724	57			

Table 5. ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2353.050	2	1176.525	6.556	.003
Within Groups	9870.675	55	179.467		
Total	12223.724	57			

4 Conclusion

Vocabulary competence is a multidimensional concept that includes both different types of vocabulary knowledge and the ability to actively utilize this knowledge in oral and written communication. The findings of the study confirmed that corpus-assisted second language vocabulary learning can expand the participants' depth of vocabulary knowledge and promote their vocabulary development; there are differences in the depth of second language vocabulary knowledge acquisition among learners with different second language proficiency level and vocabulary size. The vocabulary competence of foreign language learners is one of the most important factors affecting the success or failure of foreign language learning. Therefore, it is of theoretical and practical significance to systematically and deeply explore learning methods and tools that can effectively improve L2 learners' depth of vocabulary acquisition and develop their vocabulary ability.

Acknowledgments

This study is financed by ① Undergraduate Educational & Pedagogical Research and Reform Project of Chengdu University of Information Technology, 2022. Project number: JYJG2022054. ② 2024 Degree and Graduate Education Teaching Reform Project of Chengdu University of Information Technology, Project number: CUITGOKP202404.

References

1. Zhu Y., et al: A research on computer-assisted language learning in the age of intelligence. *Foreign Language Education*, 42(5). 51-56 (2021).

2. Qi Y. L.: The theory and practice of computer and network assisted English teaching. Zhejiang Gongshang University, Hangzhou (2020).
3. Nation, I. S. P.: Learning vocabulary in another language. Cambridge University Press, Cambridge (2001).
4. Mirosław P. & Mariusz K. (eds.): Individual Differences in computer assisted language learning research. Routledge (2023).
5. Read, J.: Validating a test to measure depth of vocabulary knowledge. In Validation in Language Assessment, ed. A. Kunnen, pp 41-60. Erlbaum, Hillsdale NJ (1998).
6. Dang, TNY & Long, X.: Online news as a resource for incidental learning of core academic words, academic formulas, and general formulas. TESOL Quarterly. 58(1). 32-62 (2024.).
7. Corder, S. P.: Error analysis and interlanguage. Oxford: Oxford University Press (1981).
8. Webb, S.: The effects of repetition on vocabulary knowledge. Applied Linguistics. 28: 46–65 (2007).
9. Liu S.: Exploring word knowledge and its acquisition patterns: An experimental study of word meaning and affix. Foreign Language Teaching and Research, 33(6): 436-441(2001).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

