

A Study on the Influence of Computer Corpus Software on College Students' English Vocabulary Learning

<https://doi.org/10.3991/ijet.v12.i08.7137>

Xiaoying Wang^(✉), Dan Zhang
Zhejiang University, Hangzhou, China
what_gonna_do@126.com

Abstract— English vocabulary and lexical collocation and use are the basic manifestations of college students' English proficiency and ability. This paper puts forward the application of computer corpus software to the college vocabulary teaching. Through the definition of probability, application prospect and theoretical explanation, this paper expatiates on the process of computer corpus serving college English vocabulary teaching. With English national corpus (BNC) as the software basis, we conducted an empirical research on vocabulary teaching. Finally, the differences between computer corpus teaching and traditional vocabulary teaching are analyzed and discussed through experimental comparison. The experimental results show that the computer corpus based collage English vocabulary practice is of larger breadth, deeper depth and higher initiative in vocabulary learning.

Keywords— computer corpus, college English vocabulary, BNC (British National Corpus), traditional vocabulary teaching

1 Introduction

Human expression is limited without grammar and impossible without vocabulary. As one of the three main components of language, vocabulary is the material in constructing language and the unit of language meaning, playing an important role in language study. Vocabulary teaching is a basic part of English teaching [1]. According to *College English Curriculum Requirements*, students should acquire a total of 4,795 words and 700 phrases, among which 2,000 are active words. Students should not only be able to comprehend the active words but be proficient in using them when expressing themselves in speaking or writing. It can be seen that the requirements of college English vocabulary are not limited to quantity, but include the deep understanding of vocabulary, so as to enhance the ability of lexical use. However, the current college English vocabulary teaching effect is unsatisfactory. A lot of time and efforts have been spared in vocabulary learning, resulting in lexical memorization in great quantity but problematic application, which reflects poor quality of vocabulary acquisition. Students put a lot of time and energy to learn English vocabulary, although the minds of a lot of words, but the application of the problem after another. There are still many problems in the "quality" of Chinese college students' vocabulary

acquisition. We should not ignore the standard of “quality” in favor of vocabulary enlargement.

English vocabulary and lexical collocation and use plays a pivotal role in English language learning, and lexical teaching is a basic part of English teaching [2-3]. As an auxiliary tool of language teaching, corpus have abundant real materials of native English users that provide reference for collage English learning and the understanding of lexical meanings. Limited by the language environment, college students have insufficient chances to communicate with native English speakers, to the disadvantage of deep understanding of vocabulary [4]. Through the introduction of corpus teaching, this paper provides an English vocabulary learning platform to help student users grasp the lexical connotations in multiple levels and dimensions. It can be instructive for teachers to lift teaching levels and to apply educational science and technologies to traditional teaching mode. Finally, with the application of the college English platform to classroom teaching, we verify the helpfulness of corpus in improving English vocabulary, lexical depth, vocabulary matching, and vocabulary application ability.

2 Computer corpus

2.1 Concept definition

The concept of computer corpus originated in Latin, meaning "body". Written and oral language materials that are processed and stored in the computer constitute a baseline library for users to study language on the computer corpus [5-6]. A corpus is a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language. According to some domestic scholars, corpus is an electronic library that stochastically collect continuous daily language and takes the advantage of texts and dialogue sections. Some scholars believe that corpus is a textual database generated after the computer processes foreign language. In our understanding, corpus is established as a randomized sampling of typically-used real language materials. It is an important basis for language research and teaching, and the basic source of teaching material compilation, grammar books and dictionaries [7].

2.2 The main types of computer corpus

In terms of its use, corpus can be divided into general corpus and professional corpus. In this paper, general corpus is our research target and application basis, represented by the English national Corpus (BNC), the American National Corpus (ANC), and Bank of English, (BOE) (Hoffmann & Lehmann, 2000).

According to language variants, corpus can be divided into native language corpus and learner corpus. The language samples collected in the native language corpus are all derived from native speakers [8-10]. In the analysis of non-native speakers or learners use language characteristics, the native language corpus is a common reference. Learner corpus is composed of non-native learners' language use samples, play-

ing important roles in observing learners' language development, language errors and foreign language teaching and research [11].

2.3 Application Prospect of Computer Corpus in English Teaching

The English knowledge in corpus is intuitive, accurate and scientific. The application of corpus in English teaching has an optimistic outlook in non-English-speaking countries [12]. It can not only provide professional guidance to English teaching and learning from the perspective of vocabulary, grammar, and article structure, but also establish a bridge of more intuitive understanding between instructors and learners with corpus data.

The traditional mode of English teaching is usually 3P teaching [13-15], in which learners access and learn English contextual materials under the guidance of teachers. The corpus-based English learning method, i.e. data-driven learning (DDL), is different from the traditional language teaching model. It changes the role of teachers and learners in English learning, by expecting learners to be more frequently-participated and active in English learning and teachers to be able to use the database to monitor learners' learning state and ability.

3 BNC software service based on computer corpus

3.1 English learning object

The English learning object of our study is College English Intensive Reading (1)" (the third edition), Shanghai Foreign Language Education Press (2006). The teaching contents cover 8 of the total 10 units in the book, with 413 words and phrases. The key words in each unit should be taught to learners in depth. They are searchable in BNC software so that learners and teachers can deepen the understanding and mastery of these new words in the process of teaching.

3.2 Experimental corpus and software

This article uses BNC, one of the most typical contemporary English corpora, as our software. It contains 100 million vocabulary and abundant oral and written texts. Figure 1 shows the user interface of BNC.

BNC learns and understands target words in the formats of list, chart, collocates, compare, keywords in context (KWIC) and so on. Figure 2 is the BNC search interface.

Now we take an English word "strategy" as an example, search it in "KWIC" and show the search results in charts. In this way, we have an intuitive understanding of the word's application in such fields as spoken language, novels, magazines, academic, non-academic, military, and newspaper. Moreover, learners are accessible to the lexical contexts that are similar to their own application context, so that gaining deeper understanding of the word (concordance). Figure 3 is the search classification chart of "strategy".



Fig. 1. BNC user interface

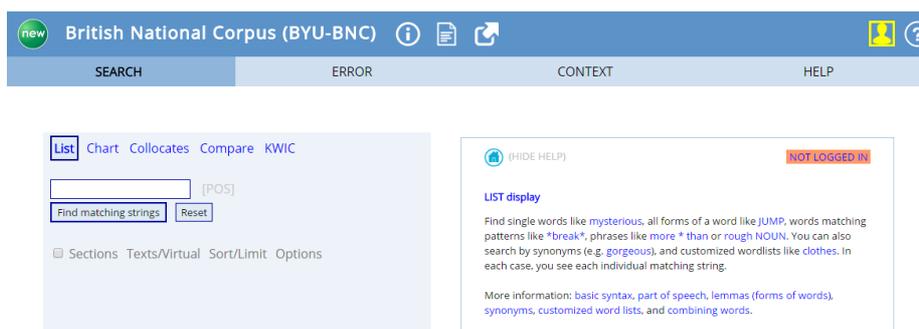


Fig. 2. BNC search interface

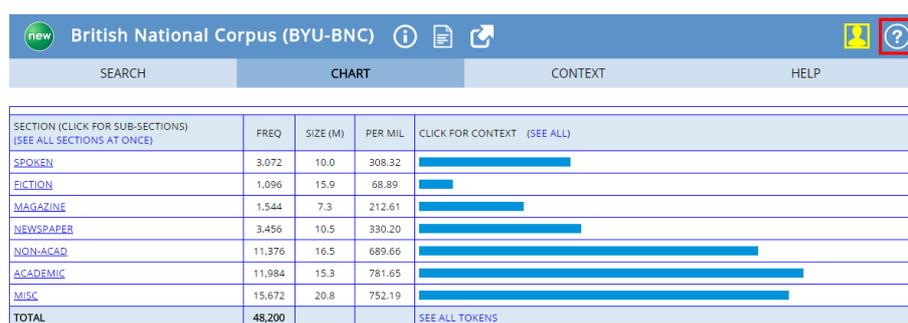


Fig. 3. Search classification chart of “strategy”

4 An Empirical Study on the Influence of Computer Corpus on College English Vocabulary Teaching

4.1 Experimental subjects

In this paper, we chose the freshman of non-English majors in two private colleges as the object of study. Objectively speaking, their initial English levels are close to each other, because of the similar scores of English achieved in college entrance examination. In order to meet the needs of experimental comparison, the fifth class is selected as the experimental group and the seventh class as a control group. The teaching lasts for a semester or a total of 18 weeks. The frequency of English vocabulary teaching is 35 minutes every fortnight.

4.2 Experimental purposes

The purpose of this study to raise the college students' levels of the understanding, memorization and application of vocabulary knowledge. The computer corpus can deepen understanding and memorization of vocabulary, lexical collocation, and contextual meaning on the one hand, and cultivate self-learning and self-exploration ability and enhance the students' ability of knowledge induction.

4.3 Specific experimental steps

For the controlled group: in the traditional 3P teaching mode, we tested the initial vocabulary level and, after 18 weeks, tested the final level of vocabulary.

For the experimental group: the computer-corpus-based data-driven learning methods were employed to teach college English vocabulary: 1. Read aloud the glossary tables, and correct the pronunciation. 2. Familiarize students with concordance example sentences. 3. Assign classroom relative to the collocation and use of vocabulary. 4. Encourage independent thinking and summarization of vocabulary use laws. 5. Make sentences with new words as a way to deepen the understanding of vocabulary.

The purpose of the post-experiment test is to verify whether the experimental hypothesis proposed in this paper is established. After the experiment, a questionnaire survey was carried out on experimental group members to test the effect of computer-corpus-based teaching method on the enlargement of English vocabulary and the improvement of independent vocabulary learning ability.

4.4 Comparative analysis of experimental data

A total of 83 students participated in the comparative teaching experiment and the English vocabulary pre-test and post-test. The test contents are mainly composed of three parts: filling in the blank, for the test of vocabulary collocation ability; multiple choices, for the test of vocabulary analysis and synonym discrimination; and word meaning, for the test of vocabulary understanding and use.

Comparison of pre-test results . We used SPSS17.0 to compare the pre-test statistical data of independent samples between the experimental group and the controlled group, and the results are listed in Table 1.

It can be seen from Table 1 that in the pre-test test, the two groups are slightly different from each other in average score, standard deviation and standard deviation of average score, indicating the similar initial level of English between testees, which meets the requirements of test prerequisites.

Comparative analysis of post-test results. As can be seen from Table 2, in the post-test, the average score (53.9070) of the controlled group is 1.1628 higher than the pre-tested one, while the experimental group has an average score of 62.0000, which is 10.4 higher than the one before the semester. Meanwhile, it is easy to find that although both their vocabulary masteries are raised to a certain degree, the average score of the experimental class is 8.093 higher than that of the control class.

Comparative analysis of pre-test and post-test scores. In order to further know the influence of computer corpus teaching on the degree of college English vocabulary, we conducted a comparative analysis of each item, as shown in Table 3.

Table 1. Pre-test independent sample statistics of the two groups

	Class	N	Average	Standard deviation	Standard deviation of average
Pretesting	Controlled group	43	52.7442	16.62484	2.53526
	Experimental group	40	51.6000	16.66443	2.47676

Table 2. Post-test independent sample statistics of the two groups

	Class	N	Average	Standard deviation	Standard deviation of average
Post-testing	Controlled group	43	53.9070	14.99018	2.28598
	Experimental group	40	62.0000	15.07247	2.38317

Table 3. Independent sample statistics of different indicators in pre-test and post-test groups

		Average	N	Standard deviation	Standard deviation of average
Contrast 1	Total points- pretesting	51.6	40	15.66443	2.47676
	Total points- posttesting	62	40	15.07247	2.38317
Contrast 2	Language collocation- pretesting	15.275	40	6.38101	1.00893
	Language collocation- posttesting	18.75	40	5.26600	0.83263
Contrast 3	Language class connection- pretesting	14.65	40	4.12342	0.65197
	Language class connection- posttesting	17.75	40	5.08265	0.80364
Contrast 4	Contextual meaning- pretesting	21.675	40	7.64396	1.20862
	Contextual meaning- posttesting	25.5	40	6.72538	1.06338

As can be seen from Table 3, the post-test score increases by respective 3.47, 3.10, and 3.82 than the pre-test score in terms of language collocation, language colligation and mastery of contextual meaning. The data results show that the experimental class

members perform much better in vocabulary scope and depth after participating in computer-corpus-based vocabulary classes, and they have attained higher-level understanding of language collocation, language colligation and contextual meaning. This phenomena verifies the positive effect of computer corpus based vocabulary learning on college English vocabulary teaching.

Questionnaire survey of students in experimental class. As our tests are conducted on unaware testees, experimental group students learns English vocabulary on the basis of computer corpus unwittingly. After the semester ended, we investigated on them with another questionnaire, in which four questions were involved: Question 1. The emotional attitude on corpus; Question 2 opinions on corpus-based teaching method and traditional teaching method; Question 3 whether the corpus is helpful in improving self-exploration ability; Question 4 whether the corpus teaching method is dependent of classroom guidance. In question 1, A means like, B is lesser like, C is unemotional; in question 2, A is better than, B is equivalent, C is just-so-so; in question 3, A is greatly improved, B is improved, C is not improved; in question 4, A is of some need, B is yes, C is no.

Through the questionnaire survey of students in the experimental class, the results of the survey questionnaire are shown in Table 4.

Table 4. The percentage of questionnaire

Question/Option	A	B	C
1	9/22.5%	24/60%	7/17.5%
2	23/57.5%	12/30%	5/12.5%
3	19/47.5%	16/40%	5/12.5%
4	25/62.5%	13/32.5%	2/5%

From the questionnaire in the table, we can see that there are 33 students (82.5% of the total number) that like or lesser like computer-corpus-based English teaching, indicating the well-received effect of this teaching method. 57.5% of students think that corpus teaching is superior to traditional teaching. In the experimental class, there are more than 87.5% of the total number of testees suggest that corpus vocabulary learning is helpful in improving self-learning initiative. In question 4, 38 students (95% of the total number) indicate that they need guidance from the teacher to complete the corpus-based English vocabulary study.

5 Conclusion

In this paper, with the British National Corpus as a software foundation, we explore the effect of computer corpus based English learning on college students in English vocabulary teaching. The research result is of great instructive meaning to improving students' English vocabulary level and laying the foundation of English knowledge learning. What is more, we conducted several experiments to verify the positive influence of computer corpus on raising university students' English vocabulary levels, drawing the following conclusions by using questionnaire survey:

1. Computer corpus based vocabulary learning method is advantageous for students to enlarge English vocabulary, to strengthen lexical scope, and to use vocabularies in contexts.
2. Compared with the traditional teaching mode, computer corpus teaching can achieve better academic results and develop student initiative in learning English.
3. Due to its unfamiliarity to students, computer corpus teaching should be used under the guidance of teachers, so as to assist students in absorbing corpus knowledge more effectively.

6 References

- [1] Catania, A. C. (1969). Editorial: on the vocabulary and the grammar of behavior. *Journal of the Experimental Analysis of Behavior*, 12(6), 845-846. <https://doi.org/10.1901/jeab.1969.12-845>
- [2] Barreras, M. (2010). How to use tales for the teaching of vocabulary and grammar in a primary education english class. *Breast Cancer Research*, 11(4), 1-11.
- [3] Tognini-Bonelli, E. (2006). Corpus linguistics at work. *Computational Linguistics*, 28(4), 583-583.. <https://doi.org/10.1162/coli.2002.28.4.583a>
- [4] Thibault, P. J., & Baldry, A. (2006). Multimodal corpus linguistics. *Supportive Care in Cancer*, 10(4), 272-280.
- [5] Oostdijk, N. (1991). Corpus linguistics and the automatic analysis of English. *Biophysical Journal*, 106(2), 240a-241a.
- [6] Hoffmann, S., & Lehmann, H. M. (2000). Collocational evidence from the British national corpus. *Bioorganic Chemistry*, 23(4), 380-396.
- [7] Stubbs, M. (2003). Words and phrases: corpus studies of lexical semantics. *Nature*, 71(4), 90-92. [https://doi.org/10.1016/S0346-251X\(03\)00016-2](https://doi.org/10.1016/S0346-251X(03)00016-2)
- [8] Rezvani, E., Eslami-Rasekh, A., & Dastjerdi, H. V. (2014). Teaching speech acts: input enhancement versus presentation, practice, and production. *Journal of Immunology*, 3(4), 58-61. <https://doi.org/10.5861/ijrse.2014.802>
- [9] Sánchez, R. C., & Pérez, A. S. (2010). Cognitive processes and the p-p-p (presentation practice production model) in contemporary elt materials. *American Journal of Hypertension*, 12(4), 188.
- [10] Lughofer, E., & Kindermann, S. (2010). Sparsefis: data-driven learning of fuzzy systems with sparsity constraints. *IEEE Transactions on Fuzzy Systems*, 18(2), 396-411. <https://doi.org/10.1109/TFUZZ.2010.2042960>
- [11] Kassab, R., & Alexandre, F. (2009). Incremental data-driven learning of a novelty detection model for one-class classification with application to high-dimensional noisy data. *Machine Learning*, 74(2), 191-234. <https://doi.org/10.1007/s10994-008-5092-4>
- [12] Carneiro, R. M. O. (2014). Teaching vocabulary: lessons from the corpus, lessons for the classroom. *Geochimica Et Cosmochimica Acta*, 8(1), 767-768. <https://doi.org/10.14393/DL15-v8n1a2014-39>
- [13] Padhi, S. (1973). Studies on the influence of enucleation of corpus luteum and administration of hormone on the induction of estrus with special reference to serum cholesterol in sub-fertile cattle. *Journal of Psychiatric Research*, 47(5), 628-35.
- [14] Mesterné Csernyik, A. K. (2013). Options in using the achievements of corpus linguistics in teaching english in a secondary education. *Astrophysical Journal*, 798(1).

- [15] Mahlberg, M. (2006). Lexical cohesion: corpus linguistic theory and its application in english language teaching. *International Journal of Corpus Linguistics*, 11(11), 363-383. <https://doi.org/10.1075/bcl.17.07mah>

7 Authors

Xiaoying Wang (corresponding author) is a Doctoral Candidate at School of International Studies of Zhejiang University, Hangzhou 310058, China. She mainly focuses on Corpus Linguistics, especially the interface between lexis and grammar, and phraseology. (what_gonna_do@126.com)

Dan Zhang is a Doctoral Candidate at School of International Studies of Zhejiang University, Hangzhou 310058, China. She mainly focuses on Corpus Linguistics, Translational Studies and Contrastive Linguistics. (newworld272@126.com)

Article submitted 10 May 2017. Published as resubmitted by the authors 18 June 2017.