

An Empirical Study on the Effects of Computer-Corpus-based Formulaic Sequences on College Students' Oral English Learning

<https://doi.org/10.3991/ijet.v12.i08.7140>

Hui Liang

Jilin Institute of Chemical Technology, Jilin, China
27268959@qq.com

Abstract—As a kind of stylized language expression, formulaic sequences are widely used in different spoken language contexts, but the study of English formulaic sequences has always been a difficulty in college students' oral English learning in China. This paper adopts the computer corpus approach and utilizes its massive storage of formulaic sequences and high-speed data retrieval capabilities to help learners with English formulaic sequence learning. On one hand, learners are able to truly and objectively understand the use of formulaic sequences in English-speaking environment through the computer corpus; on the other hand, learners can improve their memories and increase their formulaic sequence learning efficiency with the help of the structured and regularized and massive amount of use cases of formulaic sequences retrieved by the computer corpus. Finally, in the empirical study, through the comparative analysis of the test group and the control group, this paper proves that the computer-corpus-based formulaic sequences have great significance to oral English teaching.

Keywords—Formulaic sequences, Computer corpus, College students' oral English, Empirical study

1 Introduction

The development of modern information technology has a profound impact on the contents, methods and philosophies of foreign language teaching. Corpus linguistics is just a new discipline emerging after the information technology was successfully introduced into the general linguistics field, which is in turn accelerating the informatization process of general linguistics and being widely applied in foreign language teaching and research [1]. The corpus-based preliminary studies show that language is essentially composed of a number of stylized language blocks, i.e. formulaic sequences. These formulaic sequences are an important part of a natural, fluent and authentic language, reflecting three dimensions of language proficiency - complexity, accuracy, and fluency [2]. In recent years, with the rise of corpus linguistics and the rapid development of cognitive linguistics, the essence and roles of formulaic sequences have been recognized by international scholars, and the theory of formulaic

sequences has been playing an increasingly important role in foreign language learning and teaching research and has become a popular research topic in the field of second language acquisition.

Through theoretical discussion and empirical research, many scholars in China find that formulaic sequences play a positive role in improving the learners' foreign language communication ability and promoting foreign language teaching. Oral language expression constitutes a big part of communicative ability, however, it has been a traditional weakness of Chinese English learners compared with their excellent writing ability. Many Chinese scholars and educational researchers are actively seeking for a breakthrough in order to improve Chinese students' oral English and achieve reform and innovation in English teaching. The application of computer corpus in the study of formulaic semantics is just research that this paper tries to conduct to find the breakthrough. As an auxiliary means for English learning, computer corpus is recognized as being real-time, objective and accurate. In recent years, with the development of computer network technologies and mobile network technologies, computer corpus has been widely used in English learning. Now the most frequently used English corpuses are usually developed and run by English-speaking countries, which provide highly accurate interpretation and explanations for formulaic sequences. Therefore, searching and learning formulaic sequences based on computer corpus can greatly improve students' oral English performance. This paper interprets and analyzes the specific formulaic sequences through the Corpus of Contemporary American English, and applies the corpus-based formulaic sequence learning in the actual college oral English teaching cases. Through the comparative experiment, this paper analyzes and compares the experimental results and demonstrate that the computer-corpus-based formulaic sequences proposed in this paper have a positive effect on improving college students' oral English.

2 Definition and characteristics

2.1 Definition of formulaic sequences

The concept of formulaic sequence was first proposed by British linguists Alison Wray and Michael R. Perkins. It was defined from the perspective of mental representation as "a sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar [3].

Formulaic sequences, "stored as a whole in the brain" [4], are used to express information, continue and start turn-taking, indicate the attitude and ensure the smoothness of communication; therefore, in order to make language expression fluent, authentic and appropriate, it is important for both native speakers and foreign speakers to master a considerable amount of formulaic sequences.

2.2 Characteristics of computer corpus

A corpus is a database containing a large number of linguistic information that are used in real terms. By using the statistical method to carry on scientific quantitative analysis and the inductive method to carry on the rational qualitative analysis, it can reveal the most typical characteristics and actual use patterns of a natural language and reflect how to use a language in real terms. Therefore, the corpus-based approach often delivers more objective, true, reliable and available results, and thus plays a positive role in language teaching and research [5]. A corpus has the following three characteristics:

1. Scientific and objective. As corpus information is all collected from the authentic languages, it not only objectively and truly reflects the use patterns of all ethnic languages, but also adopts the scientific search method to make the test results more typical and make it easier for learners to carry out quantitative and qualitative analysis on the use environment and characteristics of formulaic sequences systematically.

2. Self-service and open. Computer corpus is like a huge electronic data repository, which is free from time and space constraints. Users can access the computer corpus via the Internet anytime, anywhere as needed. The corpus provides the learning materials on formulaic sequences to help teachers and learners learn and exchange on this independent open platform and make the data-driven English teaching possible [6].

3. Timely. Timeliness means the corpus is updated with the actual evolution of the language. Languages are evolving with the times, so are formulaic sequences [7]. This feature prompts a corpus to update its data constantly at real time. According to the statistics of the Corpus of Contemporary American English, from 1999 to 2012, the corpus needed to be updated once or twice a year to complete the expansion of 5 different texts at a growth rate of 200 million words a year. In this way, the corpus keeps it always current [8].

3 Application of computer corpus in oral formulaic sequences teaching

3.1 Case analysis of English formulaic sequences

In the popular American situation comedy “Friends”, there are many commonly used English formulaic sequences. The oral English expressions and contextual environments in “Friends” have always been like a text book for English learners around the world [9]. In this paper, we search the contexts and explanations for formulaic semantics commonly used in “Friends” in the computer corpus in order to deepen our understanding of the formulaic sequences and apply it as a teaching case in the oral teaching model [10]. In the script of “Friends”, there are up to 892411 words, many of which are formulaic sequences. According to the statistics made by other researchers, expressions which appear more than 9 times in the script are considered formulaic sequences.

Table 1 shows the formulaic sequences used by the person who tends to give and receive information. They mostly use the first person pronoun “I” and the declarative language, such as I mean, I think, I believe, and so on.

Table 1. Formulaic sequences of information exchange in the teleplay<Old friend>

Formulaic sequences	Frequency
I mean	1540
I know	1289
You know	1124
I think	984
I don't know	708
I'm gonna	690
I thought	392
...	...
How about you	9
How are you doing	9
Total	Item:183; Frequency:12921 times

In Table 1, if the formulaic sequences are extended to the native English language environment, they are also frequently used stylized phrases. If students can better understand the context and usage of these formulaic sequences, their oral English can be effectively improved [11].

3.2 Computer corpus COCA

Corpus of Contemporary American English (COCA) was released online in 2008. In recent years, it has seen rapid development. So far, there have been over 100,000 registered users and even more unregistered users in China. With simple and concise operating interfaces and a huge database that is updated constantly, COCA has gained the trust of many users [12]. Figure 1 shows the search interface of the Corpus of Contemporary American English.

The user can search relevant corpus information by entering the key words, and better understand the vocabulary by comparing charts and key words. Figure 2 shows the search results of the word “language”.

From the chart shown in the figure, users can clearly see the use frequency of the word in each section and in each time period.

3.3 Application of computer-corpus-based formulaic sequences in oral English teaching

This paper analyzes a commonly used formulaic sequence “I mean” and a less frequently used “you know what” in “Friends” with the help of the corpus and applies the result in college oral English teaching.

We take a conversation in Episode 1, Season 1 of “Friends” as an example. The scene is that Rachel is making coffee for Joey and Chandler in Monica's apartment.

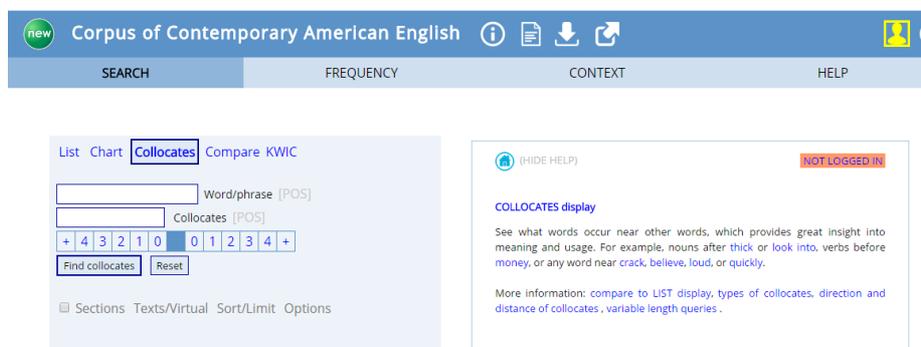


Fig. 1. Search page of COCA



Fig. 2. Search result of “language”

Rachel first says, “Isn’t that amazing?” This is a question asking for information, which completes the question function. Then Rachel says “I mean, I have never made coffee before in my entire life. Here, “I mean” serves as an intro, which is the opening words for giving information, and what follows it is the information. Together they constitutes a statement, which is a supplement to the preceding question. This supplement makes it easy for the recipient to quickly understand the information the speaker is asking for and respond in a timely manner. However, what follows “I mean” may not necessarily be information; sometimes it may also be a request for information, i.e. a question [13].

The search result in the corpus is shown in Figure 3. For example, there is a sample sentence - had nothing better to do. “the hell are they doing? I mean, that’s a lot of people’s hard-earned money,” said David. In this sample sentence, “I mean” is used in the same way as in “Friends” – as a supplement to the speaker’s statement.

Next, let’s analyze a conversation between Rachel and Joey in “Friends”. Rachel: Oh, Joey, you know what, no-one is gonna be able to tell. Joey: My mom will. In this conversation, “you know what” is a commonly used formulaic sequence. For most

Chinese students, the first thing that cross their minds when they hear this formulaic sequence will probably be “Yes” or “No” because they consider it as a question [14]. However, the actual context is that when Rachel, as the initiator of the conversation, says “Oh, Joey, you know what”, she just wants to cause the attention of the other person and then express her own opinion. The formulaic sequence “you know what” is just used as a stepping stone for the subsequent information. When hearing this expression, the other person knows that Rachel is going to start a conversation and will provide the information herself and that he does not need to respond immediately [15]. What he needs to do is be ready to give feedbacks. We use the same method to search “you know what” in COCA. As shown in Figure 4, this formulaic sequence is applied in the same way as in “Friends” – as a transition word for the subsequent information.

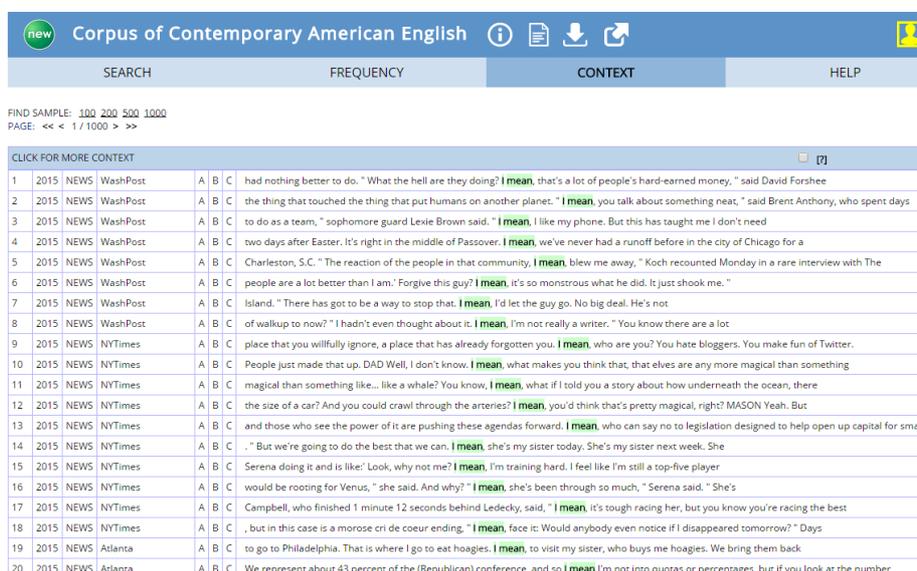


Fig. 3. Searched context of “I mean”



Fig. 4. Research context of “you know what”

Through the analysis of the formulaic sequences in both sitcom and corpus, we can see that if college students think only in the Chinese language frame and do not have deep understanding of the English formulaic sequences, they will make a lot of mistakes in their oral English expressions. This paper applies the formulaic sequences teaching in college oral English teaching based on the computer corpus in an effort to deepen college students' understanding of formulaic sequences and improve their oral English.

4 Application of computer-corpus-based formulaic sequences in college oral English teaching

This paper introduces the computer-corpus-based formulaic sequences into college oral English teaching as a teaching strategy [16]. In this paper, we select 62 freshmen of non-English majors as the subject and tries to carry out a comparative teaching experiment to study the effects of formulaic sequences teaching on the fluency and accuracy of Oral English.

4.1 Teaching experiment design

Experimental subjects. We select students of two 2010 non-English-major classes (who had similar scores in the college entrance English examination, which were all above 100) as the subjects, with 31 in each class. The subjects are divided into the test group and the control group.

Purpose of the experiment. Whether the computer-corpus-based formulaic sequences can help improve college students' oral English, and how formulaic sequences are related to the fluency, authenticity, consistency and accuracy of oral English and the amount of talks.

4.2 Detailed experimental procedures and result analysis

Pretest. First of all, we arrange the students participating in the teaching experiment to take an oral English test, which involves a one-minute self-introduction and a situational dialogue among three students. The focus of the test is to check how they use formulaic sequences. After the test, we perform independent sample analysis of the test results. The oral English pretest scores are shown in Table 2.

From Table 2, it can be seen that there is no significant difference in the oral English level between the control and the test class ($t = 0.831$, $p > 0.05$). The score of the control group is slightly higher than that of the test group.

Process of the teaching experiment. The textbook is the "New Horizon College English - Listening and Speaking". The test group, apart from receiving the routine teaching, also learns formulaic sequences based on the corpus under the guidance of teachers. With the help of the open corpus, students and teachers can communicate and exchange ideas about formulaic sequences learning without time constraints. For the test group, teachers on one hand encourages the use of formulaic sequences and

on the other hand enhance students' explicit input of formulaic sequences by using PPT and other multimedia tools in class.

The control group mainly attends the conventional communicative-approach-based listening and speaking teaching class.

Aftertest. After a semester, we arrange the students to take an aftertest in the same way. In order to better show the application effects of formulaic sequences, we add the comparison of the test group and the control group in the amount of talks, fluency, authenticity, consistency and accuracy.

From Table 3, it can be seen that after one semester of formulaic sequences learning, the oral English test scores of the test group are higher than those of the control group ($t=2.681$, $p<0.05$). It can be preliminarily concluded that the computer-corpus-based formulaic sequences teaching has a positive effect in improving college students' oral English.

Table 2. Pretesting result

Class	N	Mean	Standard deviation	T value	p value
Control class	31	83.74	4.936	0.831	0.730
Test class	31	82.69	5.003		

Table 3. Aftertesting result

Class	N	Mean	Standard deviation	t value	p value
Control class	31	83.81	3.85	2.681	0.009
Test class	31	86.47	3.96		

Table 4. Different indicators of spoken English between the test group and the control group

Class	Total words	Fluency	Authenticity	Consistency	Accuracy	Total points
Test class	17.23	17.32	17.11	17.34	17.47	86.47
Control class	16.63	16.73	16.82	16.76	16.87	83.81
Difference	5.520	5.520	1.306	5.226	5.520	109.778
Significance p	0.021	0.005	0.209	0.005	0.010	0.009

From Table 4, it can be seen that the mean values of various indicators of the test group are all higher than those of the control group. There are significant differences between the amount of talks, fluency, consistency and accuracy of the test group and those of the control group ($p<0.05$). In terms of authenticity, there is not much difference between the two groups.

The aftertest on the key indicators of college oral English further proves the positive effects of computer-corpus-based formulaic sequences teaching in improving college students' oral English. At the same time, the results show that there is a proportional relationship between the application of computer-corpus-based formulaic sequences and the amount of talks, fluency, consistency and accuracy of oral English. Due to the limitations of the English textbooks, the authenticity of oral English is not

much improved. In the future, in college English teaching, the application scope of corpus-based formulaic sequences should be enlarged to improve college students' oral English in an all-round way.

5 Conclusions

This paper innovatively proposes applying formulaic sequences in oral English teaching based on computer corpus. It first gives the definition and describes the characteristics of computer corpus and formulaic sequences, then analyzes the application of formulaic sequences in actual contexts and at last demonstrates the positive effects of formulaic sequences in improving college students' oral English through empirical research. The conclusions and the innovations of this paper are summarized as follows:

1. The computer corpus approach opens a door for the application of formulaic sequences in college oral English teaching and provides a sound data platform for the combination of formulaic sequence theory and practice
2. Formulaic sequences are essential to college oral English learning, so more efforts should be made to teach formulaic sequences in college oral English teaching.
3. The computer-corpus-based formulaic sequences provide an objective, true and real-time language environment, and thus this approach should be more widely promoted and applied.

6 Acknowledgment

“13th Five-Year” Social and Science Task of Jilin Provincial Committee of Education (JJKH20170228SK): “The Impact and Enlightenment of English Formulaic Language on Learners' Language Output”.

7 References

- [1] Gengchun, L. I. (2012). A review of the identification of formulaic sequences. *Journal of Central South University*, 13(2), 25-29.
- [2] Kersten, S. (2015). Language development in young learners: the role of formulaic language. *Journal of Bacteriology*, 135(2), 295-299. <https://doi.org/10.5040/9781474257145.ch-008>
- [3] Benito, J. F. (2009). The use of vague formulaic language by native and non-native speakers of english in synchronous one-on-one CMC. *International Journal of Impact Engineering*, 36(10-11), 1204-1209.
- [4] Kluge, E. A. (1984). The importance of foreign language education to the hospitality industry. *Journal of Nuclear Materials*, 122(1-3), 354-358.
- [5] Sampson, G. (2006). English for the computer - the susanne corpus and analytic scheme. *Computational Linguistics*, 38(1), 600-603. <https://doi.org/10.2307/415892>

- [6] Reviewer-Finegan, E. (1997). Review of "text and corpus analysis: computer-assisted studies of language and culture" by michael stubbs. blackwell publishers 1996. *Computational Linguistics*, 23(3), 487-489.
- [7] Davies, & Mark. (2014). The corpus of contemporary American English (coca). *Canadian Journal of Chemistry*, 92(1), 9-15.
- [8] L'Nyavskiy, S. (2006). A corpus-driven usage-feature analysis of the concept "fate" in modern Russian and american english languages. *Geophysical Journal International*, 166(1), 403-417.
- [9] Clopper, C. G., & Pisoni, D. B. (2006). The nationwide speech project: a new corpus of American English dialects. *Speech Communication*, 48(6), 633-644. <https://doi.org/10.1016/j.specom.2005.09.010>
- [10] Ford, M. L. (2007). Teaching strategies and standardized tests effects on educational experiences of African American English-speaking students. *Radiology*, 136(2), 465-71.
- [11] Liu, W. (2013). Role of teachers in oral English teaching. *Lecture Notes in Electrical Engineering*, 219, 13-18. https://doi.org/10.1007/978-1-4471-4853-1_2
- [12] Xie, Q. Q., Wu, S. J. (2004). Analysis of the affective factors in college oral English teaching. *Journal of Beijing Institute of Technology*.
- [13] Ghazzoul, Nahed. (2008). An empirical study of coherence and incoherence in the English academic writing of 11 arabic-speaking students. *Thin Solid Films*, 460(s1-2), 175-180.
- [14] Wieling, J., & Tump, C. (2004). An empirical study on the impact of bioanalytical method variability on estimation of pk parameters. *Chromatographia*, 59(2), S187-S191. <https://doi.org/10.1365/s10337-004-0232-x>
- [15] Zhang, X. (2011). Research on culture introduction and innovative personnel cultivation in college English teaching. *Textile Research Journal*, 85(3), 238-250.
- [16] Brenner, D. (2006). English essential for teaching assistants. *Chemical & Engineering News*, 84(8), 4-4.

8 Authors

Hui Liang is a lecturer at the School of Foreign Languages, Jilin Institute of Chemical Technology, Jilin 132022, China. She is engaged in Applied Linguistics (27268959@qq.com).

Article submitted 13 May 2017. Published as resubmitted by the author 26 June 2017.