

## Data-driven Learning and Vocabulary Teaching

Yanping Tang

School of China West Normal University, Nanchong 637002, China

### Abstract

**This paper aims to explore the use of corpus based Data-driven model in English vocabulary teaching. Firstly, it introduces the current situation of English vocabulary teaching, the origin and definition of Data-driven learning model. Then, taking coca corpus as an example, this paper presents the teaching application of using corpus to select key words, carry out word collocation and synonym discrimination. Finally, it summarizes the teaching characteristics and guidance suggestions for teachers under this mode.**

### Keywords

**Pragmatic Competence; Pragmatic Expression Ability; Pragmatic Comprehension Ability; Language Ability.**

### 1. Introduction

Vocabulary is also a necessary starting point for learning a language. Wilkins (1972) made an incisive conclusion on the importance of vocabulary in language acquisition. He said: "without grammar, many things are difficult to express; without vocabulary, nothing can be expressed". Similarly, Vermeer (1992) also pointed out the importance of vocabulary knowledge learning. He said: "Mastering vocabulary is the key to understanding and being understood. The main task of learning a new language is to master vocabulary, and grammar knowledge can not improve language ability and fluency". In China's English learning and teaching, a small vocabulary is certainly a problem. However, it is found that although many students have a large vocabulary, they still have difficulties in communication. Chu Zhenli (2006) said that how to learn vocabulary has become an important issue in English learning and teaching. Sinclair (1988) proposed in the late 1980s that foreign language teaching can start with vocabulary learning. He believes that in the past, vocabulary teaching was subordinate to grammar teaching, and the words taught by teachers were not representative and scientific. In the process of English teaching and learning, both teachers and learners have realized the importance of English vocabulary learning in the whole process of English learning. However, up to now, the most prominent problems in the process of vocabulary teaching and learning are still the time-consuming and poor effect of vocabulary learning. There are many reasons for this phenomenon, among which the lack of effective methods and means of vocabulary teaching in school English teaching is an important factor. Therefore, this paper attempts to explore new ways and advantages of vocabulary teaching combined with the booming corpus and Data-driven learning, so as to broaden vision and ideas, improve the efficiency of English vocabulary teaching and better serve English vocabulary teaching.

### 2. Current Situation of English Vocabulary Teaching

Vocabulary is the cornerstone of English language. Sufficient vocabulary is the guarantee of English listening, speaking, reading, writing, translation and other application abilities. Without enough vocabulary, these skills are water without source and wood without water for students. Without a certain vocabulary, English teaching becomes cooking without rice. Relevant

statistics show that in English learning activities, vocabulary learning puzzles 80% of students, and there are phenomena such as unable to remember and use. It is not difficult to find the root from the defects of traditional vocabulary teaching. Teachers spend a lot of time and energy on the three aspects of vocabulary sound, form and meaning. Students only mechanically and passively memorize, do not involve in deep-seated contents such as lexical chunks, collocations, pragmatic rules and contextual discourse, and fail to pay attention to helping establish the relationship between various types of vocabulary knowledge, resulting in students often only knowing the original meaning of vocabulary, Lack of a real grasp of the practical application of vocabulary, of course, it is impossible to use English naturally and accurately. The traditional "cramming" indoctrination teaching lacks students' thinking and participation activities, can not actively construct their own knowledge system, and it is difficult to internalize the knowledge taught by teachers. Therefore, Wang Rongpei (1999) believes that it is urgent to change the traditional English vocabulary teaching.

### 3. The Origin of Data-driven Learning Model

Data driven learning was proposed by Tim Johns in the early 1990s. Its purpose is to encourage learners to actively observe, summarize and summarize language facts from real corpus (Hunston, 2002:170). Johns (2002) and Chris Tribble (1900) are pioneers of assisted language teaching with corpus technology. They pioneered "Data-driven Learning" (DDL) in classroom vocabulary indexing exercises. Therefore, the Data-driven learning model is to first put forward the language problems that need to be solved. Around the problems, students can search a large number of real corpus from the corpus through retrieval tools. In addition, teachers can also print and distribute the selected corpus handouts to students. Through independent observation, analysis and discussion of the language meaning and usage in the corpus, under the guidance of teachers, Finally, the students analyze and summarize the usage rules of the language and solve the problems raised in advance. The teaching theories behind the "Data-driven Learning" model are constructivist learning theory and vocabulary grammar theory. Constructivist learning theory emphasizes student-centered active exploration, discovery and construction of knowledge. Instead of the teacher teaching directly, the students accept it passively. "Data driven learning" is to let students observe language phenomena and summarize them, so as to draw the law of practical application of language and avoid rote learning. This discovery learning is consistent with the constructivist learning view Under the "Data-driven Learning" mode, the corpus index method can well present the multiple uses of words in different contexts, effectively make up for the shortcomings of the example sentences given in the teaching materials, and comprehensively reflect the different applications of words.

### 4. Definition of Data-driven Learning

Willis (1990: 57) linked DDL with "replication activities" in teaching and believed that DDL is a classroom communication activity trying to reproduce the "real world", because real communication includes the exploration of unknown components, the collection of information and the solution of problems. Through a series of sentences with target structure extracted from a large corpus, students will try to reveal the rules of the use of target structure. In order to complete the task, learners must focus on the target language for a long time, cooperate in pairs or groups, and truly realize the discussion and negotiation of meaning. This is to define DDL as a kind of classroom teaching activity from the perspective of communicative teaching. The definition of DDL given by Johns (1991) is to use the "contextual co-occurrence" interface in the classroom to enable students to explore the laws and patterns of the target language, and design activities and exercises on the basis of contextual co-occurrence. Context co-occurrence, namely kW IC (key word in context), is the indexing program of corpus. The program uses the

high-speed retrieval function of the computer to arrange the times and specific context of the search items in one or several texts in the form of the middle of the search items in an instant, so as to observe their use characteristics and context. Compare Willis and Johns's definitions of DDL. What they have in common is to limit DDL to the classroom. However, at present, the use of DDL is not limited to classroom activities under the guidance of teachers. After mastering the use of retrieval tools, students can search in the corpus by themselves to answer the questions they encounter in learning. If students want to use the word information in their writing, but hesitate whether to use information or information, they can use some free online search websites, such as logging in to BNC website to search information online. It is found that its frequency is 38327 times, while when searching for information, it is found that its frequency is only 15 times. The ratio of 38327: 15 may prompt the student to decide to use information in his composition. The search results also provide example sentences containing the word. Students can master the usage of the word by observing these sentences. This is an example of the successful application of DDL in students' autonomous learning. It can be seen that it has gone beyond the traditional classroom learning activities and become a learning mode in which students independently carry out discovery learning and exploratory learning.

Johns emphasizes the use of "contextual co-occurrence", as if DDL depends on contextual co-occurrence. However, so far, the functions of corpus retrieval tools are far from limited to contextual co-occurrence, and many functions can be used in language learning; In addition, there are many corpora with part of speech, syntax or phonetic codes in the world.

## 5. Research on Data-driven English Vocabulary Teaching based on Corpus

### 1. Using corpus to select key words

At present, the total number of English words has exceeded 2 million. In the process of English learning, it is impossible and unnecessary to master such a large number of words. In the limited classroom time, it is very necessary to choose key words for teaching. But what are the key words and which words should be explained in detail, the teachers have a lot of randomness and subjectivity. The word frequency statistics function of the corpus provides an objective basis for our word selection. According to the statistics of Francis and Kucera (1982), in the brown corpus of more than one million words, the word frequency is the first 2000 words, accounting for about 80% of the words in the text. In other words, if English learners can master these 2000 high-frequency words first, they can basically understand the general idea of general articles. Li Xiuwen (2018) believes that high-frequency words appear frequently in corpus and have a wide range of registers, which are very useful for learners. Teachers should explain and teach them in detail in class.

### 2. Lexical collocation using corpus

Relevant researchers believe that lexical collocation is a kind of "habitual co-occurrence words", and lexical collocation is constrained by conventional phrases to a certain extent. Our students recite words in isolation, recite the Chinese interpretation of words, and combine other words according to the Chinese meaning. Chinglish appeared. In practical teaching activities, teachers can use the corpus to let students directly observe the tendency of native language users in using the vocabulary, reduce the randomness and limitations of vocabulary collocation, and avoid being affected by the negative transfer of mother tongue.

### 3. Discrimination of synonyms using corpus

The process of foreign language learning has brought a lot of confusion and confusion. By using DDL learning mode, teachers search the synonyms that need to be discriminated in the corpus, present real language examples, and organize students to observe, analyze and summarize. It is helpful for students to master the correct use of synonyms. For example, "stable" and "steady" are interpreted as "stable" and "stable" in the dictionary. Teachers retrieve the comparative

corpus of the two followed by nouns from the "compare" function option of coca corpus. The top 10 search results are shown in the screenshot. In the left column of Figure 1, the above conditional collocation of stable and condition occurs 556 times (W1), while the collocation of steady and condition occurs only 4 times. It can be seen that stable and condition can form a fixed collocation, while steady can be ignored. Similarly, in the steady column, it is matched with decline 427 times (W2), while stable is matched with stream only twice. Obviously, the frequency retrieved from the corpus can reveal the most commonly used patterns in communication. From the observation and analysis of the micro text, it can be seen that the collocation of stable and condition / system / environment mainly emphasizes the stable and unchanged state, which belongs to static adjectives. And steady, with stream / beat / pace / decline, emphasizes continuity and belongs to dynamic adjectives.

## 6. Characteristics of "Corpus Driven Learning"

Cheng Chunmei (2007) believes that the corpus driven learning model has the following three characteristics:

1) Based on corpus. Each corpus is built under the guidance of scientific database building principles. The corpus has a certain representativeness, and it is the real language used in social life. Compared with a large number of languages in some English textbooks, which are specially prepared to teach students to learn a language rule, these real corpora can undoubtedly serve as a bridge between language learning and real application, and lay a better foundation for students to use the language in practice; In addition, the codes in the code attached corpus are published only after the inspection of experts, which is scientific and rigorous, which is equivalent to adding a rich source of knowledge to students. At present, the function of corpus retrieval tools is becoming more and more powerful. "Context co-occurrence" can present a certain language knowledge point at the same time, "vocabulary" can arrange all words in the selected corpus according to the frequency of occurrence or alphabetical order, so as to provide a basis for which words should be mastered first. Other functions are not described here.

2) Learner centered. Through "problem-solving" activities, students focus on the target language materials provided by teachers for a long time, which is conducive to the acquisition of language knowledge. In the process of completing the task, students carry out exploratory learning and discovery learning, and discover language rules by themselves, rather than directly inform the rules by teachers. Students become researchers and task completors, Can better develop their learning ability. After being familiar with the learning mode of DDL, learners can also use retrieval tools to retrieve the corpus according to their own learning needs and solve the problems they encounter in the learning process. For the same language phenomenon, students can also search in different corpora to understand the regional differences and diachronic changes of language, compare the differences in use between native speakers and learners, and truly reflect the idea of autonomous learning.

3) Wide range of applications. DDL can be implemented not only in the classroom, but also in the distance education environment; Both native speakers' corpus and learners' corpus can be used; It can be either written language corpus or oral language corpus. You can study individually or in groups. Flexible mode and strong applicability. DDL puts forward new requirements for the innovation and transformation of foreign language teaching ideas and methods. Its application in foreign language teaching will not weaken the role of teachers, but also provide more creative space for teachers, which can help teachers "tailor" - design teaching contents according to students' actual needs. Fast automatic processing saves teachers the time of searching materials, makes teachers have more energy and time to devote themselves to the improvement of teaching methods, designs more creative teaching activities, and enables learners to truly master the systematic knowledge of the target language in terms

of words, usage and collocation context in their self exploration. Of course, at present, the application of corpus in teaching is not mature, and DDL is still a field to be further developed. With the popularization of corpus resources, it is hoped that more and more foreign language educators will participate in the exploration of this learning model.

## 7. Application Strategies of Corpus in English Vocabulary Teaching

Corpus based English vocabulary teaching activities can narrow the gap between English classroom and real language, consolidate students' English knowledge structure, improve students' understanding and application ability of English vocabulary, and promote the reform and development of English teaching. For example, custom The word "custom" has five different meanings in different application contexts, and each meaning is relatively different. Teachers can carry out teaching guidance by using corpus, so that students can correctly grasp the meaning and function of "custom" vocabulary in different contexts.

### 1. Lexical collocation

Teachers take corpus as an auxiliary tool for English vocabulary teaching, which can provide students with a more real context and rich teaching cases, so that students can correctly grasp the methods of vocabulary collocation, the characteristics of vocabulary use and the grammatical characteristics of vocabulary. Lexical collocation takes corpus language as the main research object. It is a method often used by vocabulary, which has a certain restrictive effect on the application of vocabulary. For example, "look forward to" means "hope, expectation". By observing the application of "look forward to" in the corpus, it is not difficult to find that "lookforwardto" is often followed by nouns or gerunds in practical application. Relevant researchers believe that lexical collocation is a kind of "habitual co-occurrence words", such as "other one", "last week", "alternate the", "and the" and other combinations. It can be seen that lexical collocation is constrained by conventional phrases to a certain extent. In order to break through the limitations of phrases and phrases, teachers should reduce the limitations and randomness of vocabulary collocation in practical teaching activities, improve students' flexibility and comprehensiveness of knowledge, and improve students' learning quality and efficiency.

### 2. Inferring the meaning of Polysemy

Lexical meaning is not only limited by concept and meaning, but also affected by the context of use. By using the retrieval function of the corpus, teachers can obtain sentence cases with high authenticity from the corpus, so that students can analyze and judge the meaning of vocabulary through their understanding and cognition of a certain context and pragmatic function, so as to improve students' imagination and judgment, and improve students' logical thinking ability and innovative thinking ability, Promote the all-round development of students.

## 8. Summary

To sum up, vocabulary is the difficulty and focus of English teaching and the basis for students to learn English knowledge. In order to improve students' English vocabulary learning ability, teachers can use corpus as English As an auxiliary tool for English teaching, it can change the traditional English vocabulary teaching mode, enrich English teaching content, expand students' vocabulary, fully mobilize students' interest in English vocabulary learning, cultivate students' autonomous learning ability, cooperative learning ability, inquiry learning ability and innovative learning ability, and enable students to connect theoretical knowledge with practical application, Pay attention to the combination of theoretical knowledge and practical application, improve the quality and efficiency of English vocabulary teaching, and promote students' overall learning Face development and lay the foundation for students' later development.



Moreover, the Data-driven model based on corpus reveals a new teaching method, which changes the teacher centered indoctrination teaching model in the past classroom, encourages students to carry out inquiry learning in the real context, and organically combines input, learners' internal ability and output, so as to make input truly understood as much as possible, Advocate student-centered autonomous learning mode. The corpus contains a large number of corpus entries. What teachers need to do is to select typical sentences that are in line with students' level and can guide students to find rules from a large number of data, let students learn and guide students to draw correct conclusions.

## References

- [1] Wang Rongpei. Exploration of English Vocabulary [M]. Shanghai: Shanghai Foreign Language Education Press, 1999.
- [2] Francis, W. N .and Kucera, H. Frequency Analysis of English Usage [M]. Boston: Houghton Mifflin Company, 1982.
- [3] Li Xiuwen. Research on Corpus-driven Foreign Language Teaching Model in Local Colleges and Universities [J]. Education modernization,2018,5(18):97-98.
- [4] Yu Yanming. Data-driven Vocabulary Teaching -- Research Teaching Based on Computer and Corpus [J]. Audio-visual Teaching of Foreign Languages,2009(2):58-62.
- [5] Willis, D. 1990 . The Lexical Syllabus .Collins: COBUILD.
- [6] Johns, T. 1991. " Should you be persuaded--Two examples of Data-driven learning materials" .ELR Journal ( New Series) 4 .
- [7] Cheng chunmei.(2007). Research on the Model of Corpus Driven Learning. Foreign Language Teaching abroad (03),58-65.
- [8] Hunston, S.Corpora in Applied Linguistics [M] .Cam-bridge : Cambridge University Press ,2002.
- [9] Johns, T. Data-driven learning:The perpetual challenge[A]. In B. Kettemann & G. Marko ( eds. ). Teaching and Learning by Doing Corpus Analysis[C]. Amsterdam: Rodopi ,2002.
- [10] Tribble, C. & G. Jones. Concordances in the Classroom [M]. London: Longman, 1990.
- [11] Wilkins, D. A. Vocabulary: linguistics in language teaching [M] London: Edward Arnold Ltd, 1972.
- [12] Zhu Zhenli. The Linguistic Theoretical Model of Lexical Units [J]. Foreign language teaching, 2006,20 (5) : 20-22.
- [13] Sinclair, J. & Renouf, A. A Lexical Syllabus for Language Learning. In Carter, R. & M. Mccathy. ( eds. ). Vocabulary and Language Teaching. London: Longman, 1988.
- [14] Vermeer, A. Exploring the second language learner lexicon [A] In L. Verhoeven and J. H. A. L. De Jong ( eds ) The Construct of Language Proficiency[C] London: Benjamins, 1992, 147-162.