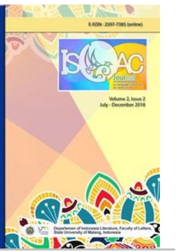
ISLLAC : Journal of Intensive Studies on Language, Literature, Art, and Culture

Volume 2 Issue 2, 2018
Journal homepage : http://journal2.um.ac.id/index.php/jisllac



A STUDY ON CORPUS-BASED EFL VOCABULARY TEACHING

Lu Hongyan[*]
(Xiamen University Tan Kah Kee College, Zhangzhou, China)

ARTICLE INFO

Keyword:
corpus linguistics,
vocabulary teaching,
word frequency,
negative language transfer

ABSTRACT

This paper introduces the development of corpus linguistics and its advantages in EFL (English as Foreigner Language) vocabulary teaching from three aspects: the application of corpus into the vocabulary teaching can greatly improve the efficiency for vocabulary teaching by providing the word frequency. And by co-presenting the actual cases of different usages and contexts of the word, the corpus can help to avoid the negative language transfer and thus enhance the learners' comprehension and acquisition of the vocabulary.

INTRODUCTION

Corpus linguistics is a relatively new linguistics study method and it has been developing rapidly since the 1980s with the development of computer science, which offers vital technical support. With its almost unparallel advantages in providing immense amounts of real corpora, and efficient and powerful search capacity, the study of corpus linguistics and its application into language teaching and learning has attracted much interest, and the importance of corpus linguistics has also been widely accepted. Many linguistic studies on how corpus linguistics can facilitate the teaching and learning activities from different levels of the language, such as corpus-based study on mistakes in collocations (Howarth, 1998; Nesselhauf, 2005; Wang, 2005; Deng 2003); corpus-based studies on high-frequency words (Ringbom, 1998; Wen Qiufang etc. 2003); corpus-based studies on modal verb (Aijmer, 2002); corpus-based studies on discourse mark (Wang Lifei & Zhu Weihua 2005); corpus-based studies on syntactic, like completive clause (Biber & Reppen, 1998); tense (Granger, 1999); corpus-based studies on discourse (Deng Yaochen, 2006), and so on.

The application of corpus linguistics in teaching can be divided into two aspects: the direct one: taking the relevant knowledge of corpus linguistics: means of developing a linguistic corpus, and applications of linguistic corpus as the teaching materials; and the indirect one: based on both corpus and computer technology, including compiling corpus-based dictionary, editing grammatical reference, textbook, developing multimedia courseware, language learning software, or evaluating or testing tools (Liang Sanyun,2005) .

Vocabulary is among the three basic elements of language and is taken as the backbone of the whole linguistic system (Ismail, Santoso, & Sunoto, 2018). As Harmer (1990) pointed out, "If the language construction is regarded as the bone of the language, then vocabulary offers it vital organs, and flesh and blood." The British linguist D. A. Wilkins once held that: without grammar, many things cannot be expressed by language, while without vocabulary, nothing could be expressed. As the importance of vocabulary has been widely realized, vocabulary teaching counts a more and more important part in the

[*] Corresponding author.
E-mail addresses: 424110710@qq.com (Lu Hongyan)

language teaching. It is natural that the study on the means of improving the efficiency of vocabulary teaching becomes the foremost of the practical linguists' attention.

Linguists turn to the corpus linguistics for the ways of improving the language teaching, which does make sense, especially for EFL vocabulary teaching. As the learners lack a real context for the application of the language, sometimes, it is difficult for the teachers to handle their teaching and for the students to acquire the usages and meaning of certain words accurately. The application of corpus linguistics into EFL vocabulary teaching can almost perfectly deal with that, for the language data in the corpus are all from natural contexts, which can help the learners to use the word accurately and properly.

CORPUS LINGUISTIC

As a research method, corpus came into use early in the eighteenth century in Europe, and all the procedures were handled artificially, which took much time and efforts. In the nineteenth century, the main application of corpus was lexicography and grammatical studies. Corpus linguistics once experienced a trough in the 1950s but began to revive in the mid-80s. The Brown corpus, built by Brown University in the 1960s, symbolized the first modern linguistic corpus, which referred to a large-scale electronic documentary corpus based on computer. Since then, different types of corpus have been built around the worldwide.

The definition of corpus linguistics

All that corpus linguistics can do is to work with a sample of the discourse. Such a sample is called corpus. As for the definition of corpus linguistics, there are many versions, which differ slightly.

(1) "Corpus linguistics studies languages on the basis of discourse." (K. Aijmer & B. Aitenberg, 1991)

(2) "Corpus linguistics is the study based on the language application in real life." (T. McEnery & A. Wilson, 1996)

(3) "Corpus linguistics can be said as a way of linguistic description or a method of testing the hypotheses of language studies." (D. Crystal, 1991)

A corpus is defined in terms of both its form and its purpose. Linguists have always used the word corpus to describe a collection of naturally occurring examples of language, consisting of anything from a few sentences to a set of written texts or tape recordings, which have been collected for linguistic study. More recently, the word "corpus" has been reserved for collections of texts (or parts of the text) that are stored and accessed electronically. Because the computer can contain and process large amounts of information, electronic corpora are usually larger than the paper-based collections that were previously used to study aspects of language. A corpus is planned, though chance may play a part in the text collection, and it is designed for some linguistic purpose. The specific purpose of the design determines the selection of texts, and the aim is other than to preserve the texts themselves because they have intrinsic value. This differentiates a corpus from a library or an electronic archive. The corpus is stored in such a way that it can be studied non-linearly, and both quantitatively and qualitatively. The purpose is not simply to access the texts and read them, which again distinguishes the corpus from the library and the archive.

Corpus linguistics, instead of studying the language itself, is a language researching method based on the corpus. It includes, on one hand, labeling the natural language date, and on the other, studying or applying the labeled language data. From a perspective of methodology, corpus linguistics can be not only applied to the study of all levels within linguistic systems but other fields beyond the linguistics as well.

Characteristics of corpus-based analysis

The word "corpus" is originally a Latin word which means "body", whose broad sense refers to the collection of discourses (the volume of the collection is not fixed). However, the modern linguistic corpus is not a simple adding or collection of the material data anymore. There are some essential characteristics of corpus-based analysis:

It is empirical analyzing the actual patterns of use in natural texts;

It utilizes a large and principled collection of natural texts, known as a "corpus" as the basis for analysis;

It makes extensive use of computers for analysis, using both automatic and interactive techniques;

It depends on both quantitative and qualitative analytical techniques.

It is the convenience and reliability that corpus as a researching method brings for the relative studies that make it so popular both within the and out of the linguistic field.

THE ADVANTAGES OF CORPUS-BASED VOCABULARY IN EFL

Compared with the traditional approaches in the EFL vocabulary teaching, a corpus-based way of vocabulary teaching has many advantages. We will mainly focus on three of these: firstly, for learners of EFL, corpus-based vocabulary teaching can help to avoid the negative transfer of the native language effectively; the statistics of the word frequency helps save the efforts in vocabulary teaching; and, the corpus can represent the natural contexts where the words are used in, which helps the learners to acquire the usages of the words accurately.

The advantages of corpus linguistics in avoiding negative transfer

In the process of EFL(English as a Foreign Language Learning), the learners will be more or less affected by his native language, which is called the transfer of the first language. The transfer from the first language can be both positive and negative. The negative transfer refers to the learners habitually adopt a certain language rule into the goal language when they encounter an occasion seemingly similar to the language familiar to them. However, the expressions of the same situation in two different languages are probably different, or even just on contrary to each other. The negative transfer disturbs our understanding of the new knowledge and communication between the two cultures, which should be avoided to the most in the process of our foreign language. The phenomenon of negative transfer is quite common in the learning of ESL vocabulary as well. For the Chinese students, their acquisition of English words is often influenced by Chinese words, whose word-construction and application are different from that of English, like:

1)* marry to me / marry with me
2)* big rain

The two common mistakes above are what Chinese students are likely to make in their English writing or expression, and they are actually the side-effects of the linguistic transfer from Chinese. The 1) mistake is made because "marry sb." in Chinese is"嫁给某人/和某人结婚". The equivalences of "给" and "和" are respectively "to" and "with" in English. Affected by the negative transfer from the Chinese, the students tend to take "marry" as an intransitive verb 2) is another mistake which is commonly seen among the mistakes made by the Chinese students in learning English. In Chinese, people use "大" to express the "heavy" rain, while in most situations, the word "大" in Chinese equals "big" in English.

When learning the vocabulary of a foreign language, learners will unconsciously turn to their first language to find the "matched" words. In fact, that is also what they get from a bilingual dictionary. For some words, on one hand, it is possible for the dictionary to contain all the usages in it; on the other, the dictionary can never explain the all the usages of the words explicitly. While taking the corpus linguistic into the vocabulary teaching, teachers can provide the students with the explicit uses of the word. The corpus is the collection of a large number of natural cases where the words are used. The abundant input of knowledge from the corpus can help to deepen the learners' all-round understanding of the word, and thus avoid the negative transfer from the first language effectively.

Frequency and ESL vocabulary teaching

The words in a corpus can be arranged according to their frequency in the corpus. Based on the data of the use frequency, the corpus can tell us the real high-frequency words in the whole vocabulary from a scientific perspective, which is rather important in the vocabulary teaching, as it decides the teaching contents in EFL vocabulary teaching. The statistics from the analysis of corpus is much more objective, and thus more reliable than our judgment from experience. According to Francis and Kucera's study of Brown Corpus, the top 1000 frequently used words count 72% of the contents in general texts, and the top 2000 cover 79.7%, and while the top 4000 take 86.8%. These data show that within the 4000words, the first half covers about 80% of the contents in the common texts, while the latter 2000 just count 6.7%, thus the first 2000 words can be regarded as the most frequently used words, which rank foremost among the list of these words that should be acquired first and they also are the basic words that the beginners should acquire. What's more, the corpus can also help to estimate what else vocabulary should be acquired besides the basic 2000 ones, which, most linguists hold, is decided by the purpose of the learners' learning English. For example, for those who are going to further their education after the basic school years, they need to acquire at least 836 academic words. To know which are the relatively high-frequent words is rather meaningful for vocabulary teaching, for it basically decides the objects, or the important points of vocabulary teaching, and can also help to arrange the sequence of the words which are going to be taught, which can make vocabulary teaching more effectively.

The advantage of teaching the words with contexts

In the corpus-based research, the corpus can be used to show all the contexts where a word occurs. From these contexts, it is then possible to identify the different meanings associated with a word. For example, the following is the display for seven occurrences of the word "deal" in LOB Corpus (LANCASTER-OSLO/BERGEN). (Because space is limited, we just list the below seven occurrences.)

1. and secret plans prepared to deal with the mass sit-down 1
2. and put one property deal through each Mr. 2
3. in particular, a good deal of concern has been 3
4. hangs a tale- and a great deal of money. Neville 4
5. where his new measures to deal with Britain's 5
6. just a matter of working a good deal harder before we really 6
7. " I'm mixed up in a deal involving millions 7

Sample "Key Word in Context" (KWIC) concord listings

The KWIC files can reveal the different meanings of words. In the above list, the surrounding contexts show three different meanings of "deal". The first and fifth entries convey a sense of handling some problem, e.g., dealing with a "mass sit-down." The second and seventh entries concern business transactions- a property deal and a deal involving millions. Finally, the remaining entries use the deal as an amount, with "good" or "great" preceding it- a good deal of concern, a great deal of money, and working a good deal harder. Because the researching value of a corpus is supported by the computer, it is relatively easy to get a list of all the occurrence of a particular word in context, and through the context, we can see all the meanings associated with the word. In fact, the complete concordance listing for "deal" shows that it has many other meanings, including:

-"to have as a topic"

(Two articles in astronomy deal in turn with stellar evolution and the determination of stellar distances)

-"a kind of treatment"

(So let us see them get a fair and square deal)

-"a type of wood"

(He went to the far end of the deal table and sat against it…)

Corpus linguistics makes it possible to identify the meanings of words by looking at their occurrences in natural contexts rather than relying on intuitions about how a word is used or on incomplete citation collections.

CONCLUSION

As vocabulary has been taken as one of the most important aspects of language, the importance of vocabulary teaching in foreign language teaching has aroused more and more attention from linguists. Corpus linguistics, based on the application of the computer technology, shows its superiority in EFL vocabulary teaching because of the great convenience in searching and the natural contexts that it provides for the target word. Despite the advantages of corpus linguistics in EFL vocabulary teaching as we have analyzed above, we should be aware of the limitations of the date that a corpus contains as well, in order to get rid of its disadvantages and make use of its advantages. Firstly, through a corpus can provide the natural contexts where a certain word is in, while it still cannot completely represent the angels of the whole textual context and social context, so the mentioned: "natural context" is still a matter of degree. Next, all that a corpus provides can be just taken as kinds of evidence, instead of information. Then, though a corpus has every reason to "boast" of the sources that it contains are of representativeness, the summarization of the data from a corpus is, in fact, a kind of deduction and instead of making statements for the language and linguistic register, all the statements in a corpus can be taken as a description of the corpus itself.

REFERENCES

Biber, D; Conrad, S; and Reppen, R. (2000). *Corpus Linguistics* [M] Beijing: Foreign Language Teaching and Research Press.

Bocheng, Z. (2006). *Corpus and Vocabulary Teaching*. Journal of Chongqing Institute of Transportation (Social Science Edition) [J], June 2006.

Guoliang, Y. (2009) *Research and Application of Corpus Linguistics* [M].Chengdu: Sichuan University Press.

Hunston, S. (2006). *Corpus in Applied Linguistics. Guided by Feng Zhiwei*[M]. Beijing: World Publishing Corporation.

Ismail, A., Santoso, A., & Sunoto, S. (2018). VOCABULARY OPTIONS OF POWER EXPERIENTIAL VALUE OF BUGIS FEMALE TEACHER IN INDONESIAN LANGUAGE SUBJECT. *ISLLAC : Journal of Intensive Studies on Language, Literature, Art, and Culture*, *2*(1), 54–61.

Sanyun, L. (2005). *Study on Corpus and Vocabulary Teaching Strategy. Computer-assisted Foreign Language Education* [J], Otc. 2005.

Shifang, L. (2007). *The Positive Effect of Corpus on Vocabulary Teaching*. Journal of Anhui University of Technology (Social Science Edition) [J], July 2007.

Teubert, W & Cermakova, A. (2009). *Corpus Linguistics: A Short Introduction. Guided by Wang Haihua*[M]. Beijing: World Publishing Corporation,

Yu, L. (2011) *College English Vocabulary Teaching Based on Corpus*. Overseas Education [J], Jan. 2011.