

Self-access Learning and Assessment of Foreign Language Vocabulary by Corpus-based Data-driven

Zhiyan Ma

Conghua College, Guangzhou Medical University, Conghua, 510925, China

Keywords: corpus; data-driven; foreign language vocabulary; self-access learning

Abstract. The application of corpus-based data-driven in teaching is more and more common, which not only builds a platform for foreign language learning, but also provides self-access learning for students. The corpus-based data-driven and self-access learning concept requires students to learn language in relatively real language environment. This paper firstly states corpus-based data-driven learning methods and concept of self-access learning, then discusses the impact of data-driven on foreign language vocabulary learning. In the end, it analyzes the application of corpus-based data-driven in foreign language vocabulary learning. I hope that this paper could give certain reference for teachers and students.

0. Introduction

It is well-known that vocabulary is the main part of language learning. Along with the education reform developing continuously, language class teachers, especially English teachers, discuss and research how to teach vocabulary from multiple perspectives. In recent years, the informationization of teaching means such as computer networks and multimedia has been widely applied to language teaching thanks to the rapid development of information age, which made the relationship between corpus-based data driven and foreign language vocabulary learning more closely. Combining with the actual situation of the current foreign language teaching, students often memorize words mechanically that has a low efficiency. If they review the words timely, many words will be forgotten in short time. Therefore, it is necessary to apply the data driven in vocabulary learning, which can improve the efficiency of vocabulary learning and enhance students' self-access ability.

1. Connotation and Characteristics of Corpus-based Data-driven

1.1 Correct understanding of the connotation of corpus-based data-driven

Data-driven learning, referred to as DDL (Data-driven Learning), is a teaching model that applies corpus in language teaching. Teachers will guide students to use computer search engine to find the course content and to do appropriate exercises in teaching program. This teaching model is connotation of data-driven learning issued by Johns and Rings, which encourages students to find vocabulary context search in corpus by computer and special software, and then to summarize using principle of words in learning process.

In 1991, Johns divided data-driven learning into three steps, namely: raised questions, material classification and conclusion. Raised questions refer to put forward main contents and tasks of language study. Material classification refers to select and classify the searched materials with the purpose of choosing sentences in line with your ability. Conclusion refers to students summarize using principle of words through independently analyze language materials guided by teachers.

1.2 Learning Characteristics of Corpus-based Data-driven

Compared with traditional vocabulary learning, there are four characteristics by using data-driven learning:

First, data-driven learning process refers to students learn knowledge by themselves. According to the basic requirements of data-driven learning, students should cultivate self-access learning by self-management, self-supervision and self-assessment. In this process, the primary responsibility of teachers is to organize and guide students to carry out self-access learning and to give appropriate explanations and to help students deepen the impression of knowledge.

Second, it is taken the authentic language input as the main content. Authentic language

environment can improve the efficiency of language learning. While data-driven could provide large quantities of linguistic data with high-quality that is derived from authentic language exchange activities. That is to say, students can learn the most authentic language by data-driven learning.

Third, it emphasizes on the exploration and discovery during learning process. Data-driven language learning asks students to explore and discover language knowledge according to their learning needs.

Fourth, data-driven learning is a bottom-up inductive learning. There are a large number of language examples in database that could help students to summarize using principle of words, to enhance learning achievement and to improve interest in learning. What's important is the knowledge summarized by them could be easier digested to achieve a true understanding.

2. Correct Understanding of Self-access Concept

Self-access concept refers to students should independently decide content, develop learning target, choose learning method and conduct a activity of self-supervision during learning process as well as assess learning outcome. In other words, it is an ability of students to learn language and vocabulary. Self-access concept is derived from constructivism theory, which considers that the learners get the knowledge through communicating with external environment and that the learning process promotes the learners to build their individual thinking activity. Or, more generally, students use original knowledge to select and process information according to their interests or needs, which sufficient emphasizes on the students' ability to use information resources for learning. The self-access learning has a huge impact on language learning and vocabulary learning, which decides the success of foreign language vocabulary learning. Therefore, cultivating self-access learning is one of main purpose of foreign language teaching. Only improves students' ability of autonomous learning can achieve the teaching goal. Under the new curriculum requirements, foreign language teaching is no longer limited by time and space, demonstrating independent learning personality and diversity. Data-driven learning is not merely platform for students' self-access learning but also a significant role in changing learning consciousness and weak autonomous learning ability.

3. Practical Application of Data-driven in Self-access Learning of Foreign language Vocabulary

3.1 Learning Lexical Collocation by Corpus-based Data-driven Search Function

Foreign language vocabulary learning is made up of quantity and quality. The former is mainly emphasized on accumulation and increased of vocabulary size. The latter focuses on capacity of words combination and phrases & collocations. The ultimate goal is to master the language skill. Of which phrases & collocations is a necessary process that is one of most predominant problems in foreign language vocabulary learning. In the traditional vocabulary teaching, teachers explain the using principle and rule of vocabulary words by virtue of their teaching experience or by means of dictionaries. However, this method is very difficult to comprehensive and detailed explanation, leading to one-sided understanding of vocabulary by students. That's why we adopt data-driven to learn words in this meeting. First, we should search all the related words matched the target vocabulary by search engine of corpus. Second, we adopt statistical approach of Z value or M value to analyze occurrence frequency of target vocabulary with other matched words, which help us to find typical collocation of target vocabulary.

Such as learning collocation of ability, we should input ability in search interface of CLCE and BNC, and then the list with matched words will be got. It was found that the words like display, perform, prove often used by Chinese students match ability to express demonstration ability is not in the list. But we can find the phrase demonstrate ability to express this meaning. For the word ability, it is a common vocabulary. Yet there is a certain error in collocation of ability due to affected by many factors. The analysis of the data-driven make the students find ignored problems, correct errors in learning process and learn the most authentic collocation. In the data-driven learning

process, students have been transformed into active explorers and could digest language knowledge found by them firmly in mind.

3.2 Differentiate and Analyze Synonym and Homonym by DDL

There are a lot of synonyms and homonyms in English, which is a major difficult point in teaching to differentiate and analyze them. In traditional vocabulary learning, teachers usually differentiate and analyze these words by virtue of language sense or dictionary, only few examples to illustrate. However, this method is out of specific language environment without in-depth explanation of subtle differences between synonyms. As you can imagine, the result has little effect. Data-driven learning could solve this problem through its high quality and large quantity of language input. There is a real and meaningful language environment in corpus. Students could find subtle differences in usage, meaning of synonyms and homonyms according to their existing knowledge structures and vocabulary building to achieve accurate use of the word. In the data-driven learning, Sketch English search procedures has a very strong comparison function for use of synonyms and homonyms, which could provide accurately discriminate.

We take preserve and conserve as example. First click the Sketch-Diff in the top of window. Second inputs preserve and conserve in lemma box. Third set relevant parameters and click Show Diff. In this time, the collocation of corresponding words in BNC will be extracted separately, ordered with usage frequency. Here we only analyze object of these two words. According to results of the query, students can distinguish the meaning of these two words based on different objects of the examples. The object of Preserve are generally abstract nouns, such as preserve integrity, preserve independence, preserve balance, which can be drawn a conclusion that this word means maintain a certain state. While the object of Conserve has a specific meaning, such as conserve energy, conserve heat, which can be drawn a conclusion that this word means protect and save. Hence students could find subtle differences between synonyms and homonyms by corpus-based data-driven and become masters of learning.

3.3 Grasp the Typical Usage in Vocabulary by DDL

In foreign language learning, there is habitual collocation between some words and other words with some characteristics to show semantic orientation, which could be used for many times to make the whole language environment with a specific semantic atmosphere, called semantic prosody which often refers to language environment. Under this circumstance, the collocation of target vocabulary usually follows certain rules to express speaker's attitude. It has certain difficulty to familiar with this method for Chinese student due to big difference on culture and language habits.

For instance, we search the word cause in data-driven and the search results will be ordered by usage frequency automatically. You can find that there are ten words like problem, damage, death, injury, harm, difficulty, confusion, disruption, distress and chaos matched with cause. The meaning of these words collocation mainly express a kind of destruction, damage and other negative things. So we can confirm that the language environment of word cause is negative. Overall, DDL can help students to strengthen the consciousness of error correction and to find the word usually used in wrong way and correct it promptly, which can help students to master the most authentic, the most typical vocabulary usage and to improve the ability to learn and use foreign language vocabulary.

4. Conclusion

Above all it is not difficult to find that data-driven teaching also slowly spread with continuous development of the corpus and information technology. In foreign language teaching, self-access learning of foreign language vocabulary by data-driven will be closely linked to the teaching of foreign language teachers, which can improve quality of foreign language teaching and foreign languages level of students. What's more, this could stimulate students' curiosity and exploring spirit and consolidate the knowledge and has very broad prospects for development.

Acknowledgments

This paper is the outcome of topics of *Research on Reform of Higher Vocational College English Studies* and excellent course construction project of *Advanced Career English* for Progress Project II of Conghua College of Guangzhou Medical University.

References

- [1] Gu Mingyuan. Study and interpretation of Opinions of the State Council on Further Increasing Fiscal Investment in Education (2010 ~ 2020) [J]. Journal of Higher Education, 2010, (07)
- [2] Liu Yuan. Practice and Theoretical Exploration of Experiential Exploration of Foreign Language Teaching [J]. Foreign Languages in China, 2011, (05)
- [3] Wang Lifei, Jiang Jinlin. A survey of the Research Focuses and Trends in Foreign Language Studies in the 2006-2010 National Social Science Fund [J]. Foreign Language Teaching and Research, 2011, (05)
- [4] Wang Lifei, Qian Juan. A Corpus-based Study on Chunk Patterns of Chinese EFL Public Speakers [J] FOREIGN LANGUAGE RESEARCH, 2009, (02).
- [5] Xin Zhiying. Systemic Functional Linguistics as an Appliable Linguistics: An Overview of Its Developments over the Last 50 Years [J] Foreign Languages in China, 2012, (03)
- [6] Qi Wenting. Application of Corpus-based Data-Driven Learning Model in College English Vocabulary Teaching and Learning [D] Lanzhou: Northwest Normal University, 2011 (05)
- [7] Wu Jiang, Jiang Lan. Discuss application of Corpus-based Data-Driven Learning in Foreign Language Vocabulary Teaching [J]. Journal of Yangtze Normal University, 2012 (08)
- [8] Ye Xiujuan. Self-access Learning of Advanced English vocabulary by Corpus-based Data-Driven [J]. Journal of Qiqihar Teachers College, 2011 (01)
- [9] Chen Ying. A Study on Corpus-based Data-Driven Method and Autonomous Learning of Second Language Vocabulary [J]. JOURNAL OF ANHUI UNIVERSITY OF SCIENCE AND TECHNOLOGY(SOCIAL SCIENCE), 2009 (12)
- [10] Chong Xingjia. Self-access Learning of Building Vocabulary for College by Corpus-based Data-Driven Method [J]. Journal of Chongqing Science and Technology University (Social Science), 2012 (02)