Research on the Innovation of English Vocabulary Teaching based on Computer Assisted Corpus

Wei Zhang
Southwest Jiaotong University Hope College, Chengdu 610400, China
*Corresponding author(E-mail: m15228122553@163.com)

Abstract
Corpus is a large scale electronic library which is constructed by the use of random sampling method and the natural appearance of continuous language. In this paper, the author analyze the innovation of English vocabulary teaching based on computer assisted corpus. In the technical condition and the information environment, it is completely feasible to establish the language database to help the learners in English learning. Corpus should be used in English teaching, it has an important practical significance, both in the perspective of teaching resources construction and the perspective of teaching reform.

Keywords: Computer-aided corpus, Data driven learning, KWIC, English vocabulary,

1. INTRODUCTION

Based on a certain linguistic principle, corpus is a large scale electronic library which is constructed by the use of random sampling method and the natural appearance of continuous language. Researches on corpus have made great progress in recent years. Corpus function is very powerful because it can quickly and accurately provide with keywords related to a large number of real corpus and in the KWIC (key word in context). It can help the analysis of key words, context and even the topic, rhetoric and other data, is very useful for the teaching of English, it can even have a great impact on the reform of teaching thinking and mode (Anna, 2009; Alfiya, 2015). As early as 90s, Johns Tim proposed the concept of corpus based data driven learning (DDL), and pointed out that learners will fundamentally become researchers in data driven mode (Fatima, 2013; González, 2015). This approach allows learners to acquire the real examples in the corpus, rather than some of the examples in the grammar books. This method enables learners to construct their own sense of language and grammar and lexical system by using an infinite number of real examples.

The majority of teachers lacks the understanding of the corpus, and lacks the necessary technical training for the application of corpus in English teaching, which makes it difficult to popularize the corpus (Susan, 2012). With the continuous improvement of the performance of the personal computer and the price declining, electronic publications and Internet English text resources have become increasingly rich and colorful, the widespread use of scanner, teachers’ personal or small group teaching can build their own teaching corpus (Meiramova, 2015; Oscar, 2015). Corpus should be used in English teaching, it has an important practical significance, both in the perspective of teaching resources construction and the perspective of teaching reform.

2. CORPUS PROFILES
2.1. Computer aided corpus

With the birth of the computer and the rapid development of the information technology, the concept of corpus has been changed greatly. The main existing form of the corpus has evolved from the original written form of the text to the present electronic text data. Now we all know that corpus has been linked to the computer. The concept of corpus is that natural continuous language is collected according to certain principles of linguistics in random sampling method. In other words, the corpus is to get the keywords of the language and use it to represent the overall study. This definition of corpus is based on the perspective of linguistics. And in the view of the morphology of the corpus, corpus refers to the computer for storing large amount of language materials and corpus retrieval software composed of a functional system. This functional system can realize the data retrieval, statistics, analysis and other kinds of applications.

Figure 2. Corpus form

2.2. The development history of corpus

According to the existence of the corpus, the development of corpus can be roughly divided into two stages.

- The first stage is the non-electronic phase: It is before 1950s. The main features of this stage are the data shows in a non-electronic form, there is no computer as an index and management tool. Most of the data are collected by the paper, so the management of the data and the index can only be done by hand. It becomes very difficult when the data collection of management, index and statistics is large for the low speed and poor efficiency. For example, in 1845, Clark began to make a manual indexing of Shakespeare's 5 plays. It took him about 16 years to complete the task.

- The second stage is the electronic corpus stage: After 1950s with the emergence of electronic computer and the gradual popularization and the rapid development of information technology, the computer has become an indispensable tool in management and indexing the corpus. The form of corpus also changes from non-electronic form to electronic data. The processing speed and efficiency of corpus have been developing rapidly.

2.3. Statistical methods of data-based analysis

Statistics and Analysis on word frequency is an important aspect of corpus research. Most corpus indexing software of the corpus of word frequency statistics can provide two frequency lists; one is arranged in alphabetical order, another according to the size of the frequency arrangement.
The most basic analytical method of the corpus is full text retrieval and word indexing. Most of the words in the corpus index software are KWIC index, which is the keyword search with context. Different corpus indexing software is not consistent with key words, Concordance KWIC index interface is divided into two parts. On the left is the vocabulary, the frequency of each is given a keyword in the library, and the keywords in the library for the percentage. On the right is the keyword for context, the span of context length can be adjusted according to needs and can be adjusted according to the word. By clicking an index row, you can look at the context of the specific text of the index line.
3. THE THEORETICAL BASIS OF CORPUS-AIDED LEARNING

3.1. Constructivist learning theory

Most scholars believe that learners should be the active seekers for stimulation; they gradually build up an understanding of the outside world. Students' knowledge is not obtained by the teachers' teaching, but by the students' active learning. Learning process is from society, family, school experience; students gradually incorporate knowledge into their own existing cognitive system, and then formed new cognitive structure, so that their cognitive level will reach a new height. Constructivism emphasizes the role of students, it requires students to change as the main body of information processing, and teachers should change as guide and facilitator. So that, teachers should adopt new teaching mode, new teaching methods and new teaching design ideas in the teaching process, and change the traditional teaching mode.

1) Emphasize the dominant role of students: It emphasizes that the design principles of teaching design, teaching evaluation and learning environment, which can give full play to the initiative of students' learning, and fully mobilize the enthusiasm of students' learning. The achievement motivation of the learners can be explained from the three aspects of the internal driving force. These three aspects are cognitive internal drive, internal drive and auxiliary drive. The learner's cognitive inner drive is not innate, but mainly obtained. In this teaching process, students can experience more fun; it is conducive to the cultivation of students' independent thinking ability and self-feedback ability, which is conducive to the construction of knowledge.

2) Emphasize the important role of situation in meaning construction: Constructivism holds that the understanding of reality and the acquisition of knowledge involve two basic processes: assimilation and adaptation. Assimilation means that the information of the external environment is incorporated into the existing cognitive structure; the cognitive structure should be modified and updated when the external information cannot be assimilated by the existing cognitive structure. Knowledge acquisition is related to a certain "situation". Only in the specific "situation", the assimilation and adaptation of these two basic construction processes can be realized smoothly, so as to complete the construction of knowledge. In the traditional classroom teaching, because of the lack of a large number of vivid practical situations, this has a negative impact on the students' knowledge construction.

3) Emphasize the role of cooperative learning in the construction of meaning: Constructivism holds that the interaction between the learner and the external environment plays a very important role in the construction of knowledge. By discussions, debates, collaborations and other learning methods, learners can exchange ideas, share wisdom and learn from each other. This process is conducive to the construction of knowledge.

4) The design of learning environment: An important difference between constructivism and student centered instructional design is the design of the learning environment. Constructivism holds that knowledge is gradually constructed by assimilation and adaptation process in the interaction between the individual and the surrounding environment. For the students, the knowledge is mainly derived from the learning environment. Without a good learning environment, the process of knowledge construction cannot be carried out smoothly, so as to influence the improvement of students' knowledge level.

3.2. Data driven learning theory

Compared with the traditional foreign language teaching, the data driven learning foreign language has the following main features:

- The main process characteristics of students' autonomous learning: At present, in the mainstream of foreign language teaching mode, teachers are still the main role of the whole teaching process, teachers control the teaching arrangement, classroom organization, teaching content and related activities. Students are regarded as the auxiliary role. In the process of data driven learning, students mainly play the role of active learners, and they are the initiators, who actively participate in the entire teaching process.
as a “white board”, which is etched and painted by teachers. This kind of teaching is very easy to be criticized for a long time. It is also extremely easy to erase the part of the students’ interest in learning, and quite easy to make learning initiative contusion. Data driven learning and the teaching model make a lot differences. Both of them emphasize students' autonomous learning, completely take the students as the center and follow students' personality characteristics. It requires students in the learning process of self-management, self-monitoring and self-assessment. The improvement of autonomous learning ability will have a positive impact on other factors of students, such as learning purpose, motivation, methods, needs, emotions and so on. All these factors ultimately achieve the purpose of promoting learning. In data driven learning environment, the role of teachers is also very important. They are the organizers, negotiators and guidance in the whole teaching process to help students in the DDL method to deepen the understanding of the knowledge and to develop their ability to learn independently. In addition, autonomous learning requires students to strengthen mutual cooperation and to jointly discover the rules of language and the use of features.

- Using real language as the main language input: Due to various factors, the current foreign language teaching is still a teacher's language output and teaching materials as the main source of information for students. Because most teachers' English ability is limited, and the level of textbook compilation is insufficient, so that it is difficult for students to lean the real language. Non authentic language input may lead to a variety of language errors. It provides learners with massive amounts of data, so that it can create a real language environment for students, then improve their language intuition, as well exercise their ability to deal with language variants.
- Emphasis on exploring and discovering the learning process: Constructivist learning theory holds that the acquisition of language knowledge is not a simple process from teachers to students, but a process of discovery and exploration by students themselves.
- Advocated bottom-up, inductive learning: The present foreign language teaching mainly manifests the characteristic of the top-down and deductive way. The teacher first explains the rules of language accurately and clearly, and then explains the students are mainly through practice to strengthen understanding and consolidation. Due to the lack of sufficient knowledge of language analysis, the students are not enough to understand the rules of grammar. In this data driven learning approach, what students have learned is not prescriptive rules of grammar, but a large number of real language, from which to sum up the rules of grammar.

4. THE APPLICATION OF CORPUS IN ENGLISH LEARNING

The development of corpus has been used for a long time, but it is mainly used in information processing, linguistics research, dictionary compilation and so on. Application in English teaching is still in the exploratory stage. But based on the technical condition and the information environment, it is completely feasible to establish the teaching language database to help the learners in the English learning.

4.1. Classification of corpus

1) Teaching materials: Teaching materials refers to the content of a variety of materials in the learner's system. The corpus of selected materials should be more familiar with the students. At present, the content of the corpus is not paid enough attention in corpus teaching application. The use of intensive reading textbooks is not sufficient. Reading materials are generally selected carefully in accordance with the syllabus requirements and teaching target. So reading materials are usually with high-quality.

2) Reading materials: Listening, speaking, and reading, writing and translating are the five basic skills in foreign language learning. In general, the requirement for “reading” is relatively high. The reading material refers to the text data collected by a wide range of channels, which is used to give the students reading practice. The choice of reading materials should be fully considered in the vocabulary, subject matter and style of the article. The vocabulary number should be consistent with the learner's current English level; there can be new words but not too many. Reading materials can play a certain role in the consolidation of the review of language knowledge, but it is not conducive to the improvement of reading ability. Some complicated English articles are difficult to read, and are easy to dampen the enthusiasm of the students reading.

3) Test corpus: Test corpus refers to a corpus of texts that are more standardized and important for learners. Because the proposition of these tests is stricter, the reliability and validity of the test questions are better, so it can be used to provide some examples of the key words in some language, which can help to study the purpose of English learning.

4) Audio visual materials: The so-called audio-visual materials mainly refer to the text information extracted from the subtitles of English movies, documentaries and other video learning materials. With the development of multimedia technology, video data has played a more and more important role in English teaching. In the course of English major, it is widely used in the course of video data, such as audio and video materials. In the teaching of non-English majors, the teaching process of non-English majors also began to use video data to help learning. Video materials such as English movies, English documentaries are lively,
interesting and rich in content. It is conducive to stimulating learners' interest in learning and beneficial for students to broaden their horizons.

4.2. Application of corpus in teaching language

By using the corpus index software and some other computer software tools together and constructing the teaching texts, we can achieve a lot of specific functions.

- KWIC index review vocabulary: By the KWIC index, learners can find out the specific content in the full text, and it will be better for students to achieve the construction and consolidation of language knowledge. KWIC index will be made into web pages to be uploaded to the relevant web site, and to help learners make vocabulary construction.

- Analysis of English reading articles: Some statistical parameters of the data are correlated with many features of the corpus, especially in the case of the various features of the corpus. From the qualitative point of view, the average length of Cet6 is larger than that Cet4, and the number of the class of the former is larger than the latter. It is in agreement with the predictions of most of us. From this example we can see that some of the statistical parameters of the corpus can help teachers and learners choose language learning materials to a certain extent.

<table>
<thead>
<tr>
<th>Table 1. Statistical parameter comparison of two corpora</th>
</tr>
</thead>
<tbody>
<tr>
<td>parameter</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>Cet 4</td>
</tr>
<tr>
<td>Cet 6</td>
</tr>
</tbody>
</table>

- Video corpus aided instruction: Auxiliary teaching with English movies and television corpus can create two vivid contexts for the learners, and constitute a video audio visual context. This is consistent with constructivist instructional design, which emphasizes the importance of context on meaning construction. We collected 150 English excellent films and documentary subtitles to construct a film corpus. All the subjects of the films are extensive, and the main parameter of this corpus is as shown in table 2.

<table>
<thead>
<tr>
<th>Table 2. Basic statistical parameters of film and television corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Words(types)</td>
</tr>
<tr>
<td>Words(tokens)</td>
</tr>
<tr>
<td>Type/token ratio</td>
</tr>
<tr>
<td>characters</td>
</tr>
</tbody>
</table>

5. CONCLUSIONS

The development of the corpus application has been for a long time, but it is mainly used in information processing, linguistics research, dictionary compilation and so on. The corpus application in linguistics is still in the exploratory stage. However, in the technical condition and the information environment, it is completely feasible to establish the language database to help the learners in English learning. This research focuses on English vocabulary based on the computer-aided corpus. English corpus can provide a rich context to learners, which is conducive to the construction of learners' language knowledge. Based on the corpus indexing software analysis, we can get the main content of the articles. With the combination of the corpus index software and other computer software, it can realize the concrete function of English words analysis. The statistical parameters of KWIC index can help teachers and learners choose language learning materials, and will help to improve the English learning.

References
