Research on the Construction of Subject Group English Corpus Based on Computer Network Informatization

Xiaozhu Zhang1,*

1School of General Education, Guangxi Vocational and Technical College, China, 530226

*Corresponding author e-mail: gzyjcs@gxzjy.com

Abstract. There are many English teaching resources in China. However, the quality of English resources can not be guaranteed. In recent years, we have been emphasizing the network of English teaching. Experts believe that the utilization rate of online English teaching resources is very high. Teachers can use the corpus resources in the network to find suitable teaching materials. This kind of teaching mode greatly enhances the students' interest[1]. The construction of subject group English corpus which is easy to use in computer network has become a hot topic in English academic circles. This paper describes the concept and attributes of corpus. Finally, this paper puts forward the future direction of corpus construction.

Keywords: Computer, Subject Group, English, Corpus

1. Introduction
In the past, English Teaching in our country paid more attention to the improvement of students' cultural achievement. Students' ability of English reading and writing has been paid attention to. With the emergence of the Internet age, people gradually realize the importance of English communication. Nowadays, the trend of international cooperation is very high. Many schools pay more and more attention to the improvement of students' oral ability. The construction of corpus of subject group English is the necessary way of English teaching reform. There is no doubt that corpus is the technical foundation of linguistic research.

In fact, many foreign language scholars have paid attention to the application of corpus in language teaching. The research of corpus in China is in its infancy. We can understand English corpus as English data resource. The construction of corpus based on computer network can improve the quality of Multimedia English teaching[2]. The application of corpus can also improve students' ability to understand English (see Figure 1). Corpus can improve the traditional concept of English Teaching in time. We briefly put forward the concrete application of English corpus. We put forward some concrete measures for the construction of Web-based English corpus.

2. Analysis of the concept and development process of English corpus

2.1. Specific concepts of corpus
Corpus refers to a collection of language resources. It is a collection of text information in the form of sampling. The establishment of corpus is a basic step in the strategy of talent cultivation of every basic language in China. Corpus is an important resource that can help people understand language deeply. The application of corpus can reveal the basic rules of daily use of English.

2.2. Basic purpose of English corpus construction
The establishment of corpus can discover the basic features of linguistics and psychology. Many experts believe that corpus is the basic reference target for subsequent research. According to the theory of sociology, the corpus with multiple resources can make people have a group learning psychology. The basic purpose of corpus construction is to arouse people's attention to English. In addition, it can also lead to an important topic of in-depth understanding of English use.

![Figure 1. The establishment of traditional corpora is similar to that of dictionaries](image)

2.3. Analysis of the development of English corpus
The progress of English corpus research abroad is very fast. Many schools in our country attach great importance to English learning. In recent years, educators attach great importance to the construction of English corpus. English corpus can truly reflect the real characteristics of English content. Now, we will find that a lot of language literature in China will describe the relevant research content of corpus. However, the research on Corpus Construction in China is still in its infancy.

2.4. Analysis of corpus application teaching
The purpose of corpus construction is to make people pay more attention to English. Therefore, the construction of English corpus can be directly used in English teaching. We can simply understand corpus as a treasure of English[3]. We can use network technology to extract a lot of English teaching resources from the corpus. Students can also use the corpus to find the English resources they want after class. Students can complete autonomous learning.

3. The analysis of the main basis for the choice of corpus of subject group

3.1. Classification of subject groups
In order to effectively study the content of corpus, we must accurately understand the system of knowledge classification. English teaching belongs to the subject of language. According to the main content of all subjects, subject groups should be mainly divided into three categories. They are language discipline, literature discipline and nature discipline. All of these three different subjects can build different corpora. Corpus construction can also reflect the main characteristics of different disciplines (see Table 1).

### Table 1. Classification of subject groups and corpus

| Subject group | Representative subjects     | Corpus                      |
|---------------|----------------------------|-----------------------------|
| Language      | English                    | Language library            |
| Literature    | Language and Literature     | Library of works            |
| Nature        | Science and engineering     | Experimental literature     |

#### 3.2. English teaching is the main goal

According to a large number of literature, we can find that a complete corpus should include a lot of language resources. We find that the purpose of linguistic corpus research is to reveal the ontological features of language[4]. The purpose of school research corpus is to find the relevant teaching path of linguistics. Generally speaking, the purpose of choosing English corpus should be close to the ultimate goal of English teaching.

#### 3.3. English resources for daily use

In the process of English learning, we pay more attention to the combination of context and sentence. In addition, English teaching also attaches great importance to oral communication. According to the relevant theories of sociolinguistics, English teaching does not only include the knowledge of the subject itself. It should also include the knowledge of human communication and natural culture. Therefore, we can think that the main content of English corpus is the knowledge of English for daily use.

#### 3.4. Sharing of English resources

Educators can use English corpus to extract high-quality English resources. Corpus should be the main course resource of English subject. The condition of sharing English resources is the normal operation of corpus[5]. Therefore, the choice of English corpus should be timely. Moreover, English Corpus must be of high quality. This method is more conducive to the real-time sharing of corpus resources.

#### 4. The key technology of subject group English corpus construction based on computer network information

##### 4.1. Application of intelligent translation technology

The purpose of English teaching is to make people understand the content of English language. The teaching of English translation is the main part of English teaching. It is necessary to build a two-way English Chinese resource base. This can facilitate people to find relevant English resources in the corpus in time. Therefore, intelligent translation technology is an important application technology of English corpus. At present, this kind of translation technology is gradually upgrading.

##### 4.2. Network dictionary technology

The content of Web-based English dictionaries is similar to that of traditional dictionaries. The main content of English corpus construction should be the main theoretical knowledge of English. Knowledge system should include vocabulary and related phrases[6]. Therefore, the application of online dictionary technology in English corpus is also essential. This can ensure that the whole English language system has the main characteristics of expansibility.

##### 4.3. Application of human-computer interaction technology
The application of network course verifies the advantages of human-computer interaction technology. Computers have become the source of knowledge. Students or teachers have become receivers of knowledge. The construction of English Corpus must be learner centered. The application of corpus should provide learners with more fun. Any operation of learners can get the systematic interaction of corpus. It can also help learners master English better.

5. Specific application of English corpus construction based on computer network

5.1. Application of multimedia corpus
We know that the teaching of English knowledge is very complicated. The structure of English knowledge is also very complex. The emergence of Multimedia Assisted English teaching improves the quality of English classroom[7]. With the gradual progress of multimedia technology, the application of multimedia corpus is that people can develop more English resources in the network. The combination of corpus and classroom teaching can more easily cultivate students' innovative thinking (see Figure 2).

5.2. Compilation of English textbooks
We can think that English corpus is a complete English textbook. The application of corpus technology will greatly affect the in-depth study of language teaching. We can combine corpus with English textbooks perfectly. The content of corpus can be used in the compilation of English textbooks. The combination of the application of English teaching materials and multimedia technology can provide targeted course teaching for students. This can provide students with a more real learning environment.

5.3. Construction of network resource platform
There are many kinds of English resources in English Corpus class. We can turn English corpus into a network resource platform. The function of resource platform can be set into two parts. The first part is the application function of teachers[8]. Teachers can extract the corresponding courseware from the resource platform. The second part is the student application function. Students can choose the English resources they want in the resource platform.

6. Analysis of the future development direction of computer-based English corpus
6.1. Expansion of English language and culture
Today's corpus construction mainly includes daily teaching knowledge and oral knowledge. These contents can only help students master basic English communicative competence. Many students can not understand the charm of language and culture[9]. The future development direction of English corpus can be the process of English language and culture. Educators can integrate the history and culture of foreign language development into the corpus.

6.2. Expansion of English dialect culture
In fact, we learn English in two ways. The first language is American English. The second language is British English. They are all dialects of the English language. Due to different accents, the narrative forms of English in many countries have changed. We can often hear English with Indian accent. Therefore, the expansion of English dialect culture is also the main development direction of corpus in the future.

6.3. Application of unified corpus in different regions
The process of English corpus research in China is very slow. There are many kinds of online English corpora. According to the corpus construction of subject group, we have been emphasizing the construction of unified corpus in different regions[10]. This can avoid the loss of English resources. This can effectively ensure the integration of English resources. In addition, many schools can also get high-level English curriculum resources.

7. Conclusion
With the help of computer network, the construction of English corpus resources is making rapid progress. The author thinks that the development space of corpus is infinite. In the future, the construction of English corpus will be perfectly integrated with English education.

Acknowledgments
On English-Chinese Translation of Hedges in Abstracts of Academic Journals Based on Corpus

References
[1] Furuta Umesaki A. A Course in English Corpus Linguistics (1) : Using COBUILDdirect and WordSmith [J]. Tezukayama Review, 2002, 93:67-102.
[2] Olohan M. How frequent are the contractions: A study of contracted forms in the Translational English Corpus [J]. Target, 2003, 15(1):59-89.
[3] Mahlberg M, Ebrary I. English general nouns : a corpus theoretical approach [J]. 2005.
[4] Sinha R M K. Mining Complex Predicates In Hindi Using A Parallel Hindi-English Corpus [C]// AcI International Joint Conference in natural Language Processing. 2010.
[5] Yang H. Chinese learner English corpus (CLEC) [J]. 2002.
[6] Olohan M. Comparable Corpora in Translation Research: Overview of Recent Analyses Using the Translational English Corpus [J]. Olohan Maeve, 2002.
[7] Kuzmenko E. Russian Error-Annotated Learner English Corpus: a Tool for Computer-Assisted Language Learning [J]. Proceedings of the third workshop on NLP for computer-assisted language learning at SLTC 2014, Uppsala University, 2014.
[8] Balabel M, Hamed I, Abdennadher S, et al. Cairo Student Code-Switch (CSCS) Corpus: An Annotated Egyptian Arabic-English Corpus [C]// The 12th Conference on Language Resources and Evaluation (LREC 2020). 2020.
[9] Ha M. Exploring Academic Vocabulary in Advertising English Corpus [J]. Studies in Linguistics, 2020, 57:223-243.
[10] Farahani M. Metadiscourse in Academic English Texts: A Corpus-driven probe into British Academic Written English Corpus [J]. Studies About Languages, 2019(34):56-73.