An English Corpus Query Model based on Pattern Recognition Technology

DAPENG WANG

ABSTRACT

Corpus has become an indispensable basic resource for linguistic research and language engineering. It has started early in the study of English corpus, and has formed a perfect corpus research standard. According to the situation of our country and the English learning in China on the basis of foreign experience of practical English corpus construction, corpus linguistics, corpus based translation studies has carried out extensive research and achieved fruitful results. The domestic and foreign research present situation and trend of the construction of English corpus based, we should highlight the interdisciplinary of a higher level in the construction of English Corpus in the future, pay more attention to the construction of corpus standardization, multi modal and intelligent. Bilingual teaching is not only the demand of higher education in China and international standards, but also the inevitable trend of training international professionals. It is also the focus and focus of the current educational reform in higher education. In this paper, using pattern recognition technology, we analyze the choice of textbook and lectures in the bilingual teaching organization, teaching language, teaching forms, which has some reference for the undergraduate bilingual teaching reform and research.

KEYWORDS

International Conference, International Conference, International Conference.

INTRODUCTION

With the rapid internationalization of English, the localization of English in China is also deepening. The academic circles began to have contact with the theory of world English China English is discussed, many scholars have demonstrated the fact Chinese English exists, discusses the theoretical significance, the concept of Chinese-English teaching enlightenment and social value. Chinese English is not only beneficial to the dissemination of Chinese civilization and culture, but also to make up for the shortcoming of the small scope of Chinese language in the world. It also has the function of condensing the members of societies and bears important social significance. In recent years, describe the research characteristic in all aspects of the English language China localization method can not provide use for just unfolding, convincing data in the paper but also to explore the deep-seated reasons of the phenomenon. The vigorous development of corpus provides a powerful tool for the systematic study of Chinese english. The international academic community has established a corpus of English varieties of various countries, but it does not include the most English variants of Chinese.

Dapeng Wang, Teaching and Research Institute of Foreign Languages, Bohai University, Jinzhou 121013, China
Study of language corpus was further developed in the nineteenth Century based on the people, try to use similar to the natural science research methods to collect real language data, description and discussion, reconstruction between language genealogy. In the early twentieth Century, influenced by positivism and behaviorism, the European tradition was inherited by American structuralist linguistics, and empiricism gradually took the lead in language studies. Empiricism holds that the main source of linguistic research is corpus, and the large collection of language instances that appear naturally in a language is not only necessary but also adequate for the study of the language.

With the help of electronic corpus is the storage and processing power, large computer and computer rapid annotation, tool storage, endowed the massive natural language materials processing, automatic retrieval, indexing and statistical analysis for large, representative, has a certain structure of the language database. Extensive corpus research includes not only the construction of corpus, the exploration of methodology, the research of specific application, but also the development and implementation of corpus analysis techniques and tools.

In the course of teaching, the teaching or learning of a course using or using some other language other than the mother tongue. The actual meaning of bilingual by countries and regions are different, so we China for example, bilingual teaching mainly refers to the Chinese, a foreign language as the main language of classroom teaching, bilingual teaching requirements of English knowledge, but not absolutely excluded Chinese. The main purpose of bilingual teaching and general course teaching is the same, namely, bilingual teaching focuses on teaching and bilingual, bilingual teaching is not only to learn a foreign language, but with foreign language learning, students should focus on master professional knowledge, in the mastery of knowledge and professional courses, the curriculum of English nouns in concept and expression habits.

THE PROPOSED METHODOLOGY

Bilingual teaching model of pattern recognition.

There is no essential difference between the teaching mode of bilingual teaching mode and common course, the main difference is reflected in the language teaching and the proportion of writing, English and professional basis which mainly depends on the nature of the course and students. The students are familiar with the contents of English and the ability to accept relatively high, should use all or most of English teaching, if students are beginners or professional English level is low, should reduce the proportion of English teaching.

The pattern recognition course is an elective course for senior undergraduates majoring in electronic information. Students are beginners in the field of pattern recognition, and as senior undergraduates, they have a certain ability to accept English. On the other hand, pattern recognition as a young discipline, the Chinese translation of many professional terms is no unified consensus, so in practice, can be used by English writing, to help students master the correct concept of unity, and explain the process by the way of the combination of Chinese and English, focusing on noun the concept, to explain, to help students understand different Chinese translation to strengthen the understanding of basic concepts, basic principle.
The use of multimedia and networking.

In order to deepen the students' understanding, improve students' interest in learning, the introduction of multimedia classroom teaching, make full use of multimedia courseware for teaching content description and image display, the figurewords sound like one, make the teaching and learning activities become rich and colorful, blending knowledge learning in lively and vivid image, so as to stimulate the students' interest in learning, and become bitter music school, and promote the development of students' thinking, enrich the students' imagination. In practice, we apply the application examples of pattern recognition to the teaching contents, which make students understand the method more intuitively and understand the principles more deeply.

In order to enrich the students' autonomous learning, providing learning resources for students a variety of pattern recognition network, including online textbook teaching software and teaching resources, teaching materials and related online tutorial and academic forum, for students from various angles to enhance classroom teaching content understanding.

Frontier knowledge.

Pattern recognition is a hot research field in current intelligent information, will be interspersed with the latest progress of pattern recognition, introduced to the students, so that students understand the knowledge of pattern recognition of the subject and the latest research trends, to further develop students' interest in learning, improve their active learning motivation. In practice, explain the combination of basic concept, the basic principle of the method, with pattern recognition related to recent advances in multimedia mode show students, make students understand the bilingual teaching pattern recognition of the necessity and importance of, and increase the power of active learning.

Common language learning courseware.

Based on the Internet of voice learning courseware on the Internet is very large, simple, such as words, online broadcast, listening test. These learning software solves the problem that students are the subject of language learning. As back word software, students can not only choose the extensive dictionary to recite, such as grade four and grade six English dictionary dictionary, the TOEL dictionary, the GRE dictionary, English dictionary, also provides some English reading novels. In the course of reading, unknown words can be inserted into the new vocabulary created by the student using a simple mouse click operation. For the new words in this book, the Ebbinghaus forgetting curve method is adopted. The students are required to review in the prescribed time, and the grades of the words are graded.

Language courseware based on pattern recognition.

The courseware mentioned above can promote students to master basic English learning, but it can not help them to apply these knowledge well. If there is a software that can monitor students' learning at any time as a teacher, and correct the pronunciation of students and determine whether students answer questions
accurately, it will definitely improve the students' learning efficiency. Pattern recognition based language courseware can be a good solution to this problem. Based on pattern recognition language courseware, its core function is to use artificial intelligence technology to identify students' pronunciation, such as phonetic symbols and pronunciation of words, pronunciation of sentences, pronunciation of paragraphs, real-time dialogue and so on. Actually, speech recognition technology has been widely used.

Online broadcasting allows students to more easily access to the latest economic, cultural, political and social knowledge, so that students can quickly integrate into social life. In English teaching, sometimes it is difficult for teachers to distinguish the pronunciation rules of American English and British English, even if they can find relevant information on the Internet, it is very difficult to really understand. With the aid of online broadcasting, this problem can be easily solved. If we can understand the language broadcast, the language application must be able to reach new heights. Many online sites not only provide audio data files, some also provide explanations, and even provide English videos with Chinese translations.

If we can combine face recognition and even behavior recognition, language teaching software will be able to integrate practical problems into language teaching. Because face recognition and behavior recognition need to deal with a large amount of data, plus the difficulty of the technology itself, it limits its application and development in this field. With the improvement of hardware technology and recognition algorithm, face recognition and behavior recognition will become possible. Using virtual reality technology, the computer can recognize accurate expression according to the facial expression, voice and behavior of the learners.

Rosetta Stone is a leader in the field of language teaching, it is available in 30 languages including English, French, Chinese and other rare languages, learning courses. Speech and other different learning methods, using the technology of computer network teaching software, the use of multimedia technology and artificial recognition technology to simulate the natural language like learning environment and learning the mother tongue. In this process, learners do not have to switch to the familiar phonetic environment, and use the new knowledge to deal with the problems in the learning process directly. In this way, it can help learners achieve the goal of learning a second language at an unimaginable rate. In the learning process, image, animation, sound and text forever together, learning by watching pictures or video, listen to the standard pronunciation and view the text prompts to language learning, to learn foreign languages as naturally as the native language learning.

Trends in English Corpus.

With the systematic corpus of the progress of the research, design and construction planning, the construction of the corpus corpus, corpus collection and processing management system will realize the standardization, an important guarantee for the further development of the corpus. The specification of corpus construction is not only convenient for users to use, but also can realize the corpus sharing to a higher degree and realize the integration of different corpora. It provides the condition for the comparative research of different corpora.
With the continuous development of computer storage technology, retrieval technology and data processing technology, storage capacity is increasing and the price is low, all kinds of data retrieval and processing of massive data analysis software tools make it easier. Compared with the traditional single modal corpus, multimodal corpus can collect oral and written Chinese, and also collect many communication factors outside the text. The development of multimodal corpus will further expand the scope of corpus research.

In recent years, with the extensive application of Internet technology, Internet technology has penetrated into every field of people's lives, the network language has begun to exert a great influence on the evolution of language, similar to e-mail, blogs, chat rooms, discussion forum and other related data can only be collected after the establishment of the Web corpus can. At the same time, the network can realize a higher degree of corpus building automation, and the selection of corpus will no longer be done by manpower, and can be automatically completed by computer through the internet.

Compared with the trend of large corpus, it has a trend towards small special corpus. These two aspects are not contradictory, mainly depending on the purpose of development. If the purpose is to explore the universal phenomenon of a language, corpus researchers will tend to build super large general corpora such as BNC. However, with the corpus of research in various fields in depth, although the General Corpus of huge capacity, still unable to meet the needs of various professional fields, such as language acquisition, pragmatic style, linguistic research, which requires researchers to establish their own special corpus in a particular field of study. It shows that it is a big trend to develop more small corpus while developing a large and general corpus.

CONCLUSION

With the development of computer hardware technology and software technology to improve the application bottleneck, pattern recognition technology in language education is a breakthrough, will play a more important role in language education, and this function is unable to achieve the traditional language education, it makes students become the main body of language learning. Use of virtual reality technology, software system using pattern recognition technology, can provide a virtual reality environment, the learning environment, can feel the real life in the real scene may encounter, so as to master and use the learned language knowledge. The construction of Chinese English corpus, so that we can use more rigorous statistical methods, analysis of the characteristics of China various aspects of the English language, and understand the Chinese using English law.

REFERENCES

1. Zhang, H., Li, J., Huang, Y. and Zhang, L., 2014. A nonlocal weighted joint sparse representation classification method for hyperspectral imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7 (6), pp.2056-2065.
2. Zhang, S., Wang, H. and Huang, W., 2017. Two-stage plant species recognition by local mean clustering and Weighted sparse representation classification. Cluster Computing, pp.1-9.
3. Chen, Y., Nasrabadi, N.M. and Tran, T.D., 2013. Hyperspectral image classification via kernel sparse representation. IEEE Transactions on Geoscience and Remote Sensing, 51(1), pp.217-231.
4. Wang, H. and Wang, J., 2014, November. An effective image representation method using kernel classification. In Tools with Artificial Intelligence (ICTAI), 2014 IEEE 26th International Conference on (pp. 853-858). IEEE.

5. Cao, J., Zhang, K., Luo, M., Yin, C. and Lai, X., 2016. Extreme learning machine and adaptive sparse representation for image classification. Neural networks, 81, pp.91-102.

6. Li, J., Zhang, H. and Zhang, L., 2015. Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 53 (10), pp.5338-5351.

7. Hu, W., Li, W., Zhang, X. and Maybank, S., 2015. Single and multiple object tracking using a multi-feature joint sparse representation. IEEE transactions on pattern analysis and machine intelligence, 37 (4), pp.816-833.

8. Qian, Y., Ye, M. and Zhou, J., 2013. Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features. IEEE Transactions on Geoscience and Remote Sensing, 51 (4), pp.2276-2291.