Development and Application of Mechanical English Corpus

Rui Han, Yanlin Yin, Xu Jiang

Heilongjiang University of Technology, Jixi, Heilongjiang, 158100

Keywords: Mechanical English Corpus; Application Study; Development Status

Abstract: In the field of modern linguistics research, field linguistics has replaced the general linguistics as the mainstream of linguistics research. The corpus has naturally become an indispensable basic resource for linguistic research and language engineering. On the basis of learning from foreign experience, China has carried out extensive research and achieved fruitful results in combination with China's national conditions and English learners in English corpus construction, corpus linguistics, and corpus translation studies. Based on the research on the status quo and trend of English corpus construction at home and abroad, it is necessary to highlight its interdisciplinary nature in the future construction of English corpus, pay more attention to the establishment and popularization of corpus construction norms, and multi-modality in corpus. In addition to focusing on the construction of large-scale corpora, it is required to also increase the construction of small-scale corpus, and pay more attention to the follow-up management and maintenance of the corpus while paying attention to the construction of the corpus.

1. Introduction

The corpus reflects real language data. It is a new type of research idea to analyze the use of language by analyzing a large number of linguistic facts. The corpus has created a wealth of linguistic materials for the study of English language, covering all aspects and areas of the English language. With the corpus of the corpus, the application of the corpus in the English language has updated the ideological concept and teaching methods of English teaching, and can meet the corpus required by students and teachers in a timely manner. Foreign countries have paid great attention to the application of corpora in English teaching, and China proposed in the first Chinese corpus linguistics and English teaching research conference held by the Foreign Affairs College of South China Normal University in 2004 that the future will focus on the application of English corpus to English teaching. Basic theory, examples of teaching materials and other research. In order to better understand the development status and trends of Chinese corpus, the author analyzes the development of English corpus in the past ten years, and believes that the effective use of network resources for the construction and application of English corpus will become a prominent problem during the 12th Five-Year Plan period.

2. Status Analysis

From the induction and analysis of the development of English corpus in the past ten years, we can see that people pay more and more attention to the research in this field. The real analysis content and a large amount of data of the corpus attract convenient people to use it. Especially in recent years, the research papers on the application of corpus to English teaching are increasing, and there are achievements in all aspects of English teaching, and it is obviously growing. In the past two years, the corpus has been applied in a wider range and more extensively.

First of all, the corpus can be directly applied to English teaching. For example, explaining the corpus knowledge, explaining the ways and means of exploring the corpus and how to use the corpus data for teaching are all direct applications of the corpus. Secondly, the corpus can be used indirectly through the compilation of dictionaries, reference grammar textbooks, multimedia corpus using computer technology for software development, etc. Professor He Anping has analyzed and summarized the actual combination of corpus and English teaching, which will be reflected in the

following three aspects: First, the combination with teaching resources, that is, when writing the textbook outline, the corpus information is included as the focus of the textbook content; The combination of teaching, that is, the use of resources in the corpus for vocabulary and grammar learning when selecting reading materials; the third is the combination with teaching research. At present, although the English corpus in China is reflected in the field of teaching and research, it still ignores the construction of teaching resources. Therefore, the English corpus still has broad development space and utilization value for the construction of English teaching resources.

In contrast to the trend of corpus large-scale, it also has a tendency to develop into small specialized corpora. These two aspects are not contradictory and depend mainly on the purpose of development. If the goal is to explore the general phenomenon of a language, then corpus researchers will tend to build a very large general corpus such as BNC. However, with the deep research of corpora in various fields, the general corpus, although large in capacity, can not meet the needs of research in various professional fields, such as language acquisition, pragmatics, legal linguistics, etc., which requires researchers to establish themselves. A specialized corpus for research in a particular field. This shows that while developing a large-scale general corpus, it will be a major trend to develop more small specialized corpora.

3. Three Trends in English Corpus

This refers to the genre or theme of specialized English corpus corpus is more specific, and the size of the corpus is correspondingly smaller. The former reflects the gradual deepening of the researchers' understanding of specific discourse communities, so the corpus is further derived from the same subject matter (such as science, medicine, business, philosophy) or similar text types and genres (such as academic papers, research bulletins, book chapters). Specific to the combination of the two (such as medical academic papers, natural science lectures) and related learner corpus (such as the Indiana Business Learner Corpus hosted by Ulla Connor et al. IBLC). The latter is reflected in the representation of the corpus and the manual labeling of the corpus, mainly because the tagged special-purpose English corpus can play a greater role, but usually due to human and material constraints, we need to be in the corpus. A balance between scale and representation. Bowker and Pearson believe that as long as the corpus is carefully designed, a corpus of as few as a few thousand words and as high as hundreds of thousands of words contributes to the study of specialized languages. The key is to ensure the corpus. Openness (ie, sustainable expansion or upgrade), sufficient corpus data, authorship of corpus, and clarification of the corpus (publishing) time span, which is closely related to the research purpose of the corpus project; such as Taavitsainen, University of Helsinki, Finland And Pahta, the Early Medical English Writing project, in order to examine the characteristics of early medical academic discourse, collected corpus data in the 1550 years as the node of the evolution of medieval and modern thinking.

A general-purpose corpus such as BNC used to be used as a reference corpus in comparison with a specialized English corpus study, but recent trends have taken into account the increasingly prominent world common language status of English, with a focus on using the "international" corpus or the English variant corpus. Of course, there are also practical reasons, that is, the relatively localized corpus is easier to collect, and researchers and classroom teachers are more familiar with the socio-cultural context of the corpus, thus enabling The ethnographic dimension that was previously lacking in corpus research has been strengthened. Professor Ken Hyland of the University of London in the United Kingdom has a paradigmic influence in this area; he pointed out that "although the results of the corpus analysis are longer than the [student]'s awareness of language use and telling us how the author is writing, it is limited to this. The danger of reifying conventions in explaining phenomena." His series of meta-discourse means used expert interviews (including discourse-based interview methods) to obtain the knowledge and practice of text constructors themselves in the process of reading and writing. This ensures the cross-check of the research methods and the reliability of the research conclusions.

Language is not the only symbolic resource system of human communication. Visual, color and space are important options for meaning construction. Therefore, the corpus of multimodal text

corpus is printed text with webpage information, hypertext of webpage, and dynamic texts such as TV commercials, documentaries, lectures, or business negotiations. The corpus-based text multimodal analysis can realize the use of computer to analyze multimodal text data and study the concrete meaning and meaning construction process of its text. Professor Anthony Baldry and his colleagues in Italy developed the theory of metafunction based on system function grammar. The Mutlimodal Corpus Authoring Systemcuts multimodal text (such as movies) into phases and subphases. The former refers to the time and time. The unit of meaning construction composed of multiple resources (action, gaze, gesture, language, sound, etc.) forms conceptual meaning, interpersonal meaning and textual meaning, and the longer stage is generally related to specific activities (such as walking, talking Sub-stage of sleeping, driving, driving.

4. Suggestions for the Construction of English Corpus

The corpus has become an indispensable basic resource for theoretical research, applied research and language engineering of contemporary linguistics. Foreign countries have started early in the research of English corpus, and have formed a relatively complete corpus development and research norms. Over the past 30 years, various corpora have been established, and various corpus related software has been successfully developed. Based on the experience of foreign countries, China has carried out extensive research and achieved fruitful results in the aspects of English corpus construction, corpus linguistics, corpus translation studies, etc. in combination with China's national conditions and English learners. In view of the significance of corpus for linguistic research, China's corpus related research will certainly further advance in depth. In the future corpus construction process, we need to pay more attention to the following points: (1) Pay special attention to the interdisciplinary integration in the corpus research process, and closely combine traditional linguistics, computational linguistics, computer science, lexicography and other disciplines. Interdisciplinary study of corpus research; (2) Formulate a unified corpus construction norm, promote the co-construction and sharing of corpus resources, improve the efficiency of English corpus use, and avoid waste of resources caused by repeated construction; (3) Put more energy into In the study of multimodal corpus, more attention is paid to non-linguistic factors other than text in the process of corpus processing, from lexical, syntactic, textual, nonverbal communication, and linguistic features (such as sound, tone, intonation, speech rate, etc.). Build a corpus with multiple dimension dimensions such as body features (gestures, expressions, movements, movements) and non-physical features (environmental surroundings); (4) Develop more intelligent dynamic monitoring functions of the corpus, so that the corpus has automatic monitoring and timely Updated ability to automatically discover new knowledge, proactively report new knowledge, and automatically update corpus (5) Develop more small specialized corpora to meet the specific research needs of more targeted service researchers, popularize corpus construction techniques and use techniques, and improve the efficiency of corpus use; (6) Pay attention to the later maintenance and management of corpus, prevent The corpus construction work has a phenomenon of anticlimax.

5. Conclusion

The purpose-use English corpus study is not the end point, but the starting point. It first assists the researchers to clarify the need to further analyze and explain the specific recurrence patterns of special-purpose English. According to this, the researchers may interpret the complexes through various qualitative or quantitative methods. Regular. It should be pointed out that the design of such corpus and the data analysis of the corpus collection to the corpus are generally carried out by the end users of the corpus (ie, the subject researchers), so the corpus size is much smaller than the general corpus, which is also There are more stringent requirements for the representation of corpus.

Acknowledgements

References

[1] Hu Kaibao, Tao Qing. A Study of the Meaning of Language and Its Motivation in Chinese-English Translation[J].Journal of PLA University of Foreign Languages,2009(04).

[2] Ren Wen, Yang Ping. Towards internationalization: the status quo and trend of the development of Chinese interpreting research [J]. Chinese Translation, 2011 (01).

[3] Wang Longzhen, He Anping. A Summary of the First Chinese Corpus Linguistics and English Education Teaching Seminar [J]. Foreign Languages (Journal of Shanghai International Studies University), 2005(03).

[4] TAN Aihua, FENG Chuanyuan. Research on the vocabulary teaching method based on corpus linguistics[J].Journal of Southwest University for Nationalities,2008(S1).

[5] Xu Hongli. Analysis of the Research and Application Status of English Teaching Based on Corpus [J].Journal of Huaibei Vocational and Technical College,2006(04).