

Construction and Application of Bilingual Parallel Corpus for English Translation of Chinese Classics

Gou Xiuli*

College of Foreign Languages, Bohai University, Jinzhou, China
1029915835@qq.com

*Corresponding author

Abstract: Chinese classics show the cultural characteristics of China's multi-ethnic groups, and are the common spiritual wealth of all mankind. They not only have important ideological value, but also contain rich cultural information. The translation of Chinese classics into English is not only an important part of the cultural exchange between China and the West, but also an important way to spread Chinese excellent culture. Corpora provide a large amount of real corpus for translation, and it is very important to build a large-scale parallel corpus with rich data, accuracy and reliability for a high-quality machine translation system. In view of the many problems faced by the translation of Chinese classics into English, such as translator quality, translation level and translation strategies, this paper, from the perspective of translators and disseminators, follows the basic principles of corpus construction, designs a corpus construction scheme for the translation of Chinese classics into English, builds the corpus according to the construction process and conducts application analysis, so as to promote the introduction of Chinese classics to the world.

Keywords: Chinese Classics; English Translation; Bilingual Parallel Corpus; Construction and Application

1. Introduction

In the context of globalization, cultural contacts and exchanges between countries have changed from spontaneous to conscious, from passive to active, from one-way to two-way, and from hegemony to equality. The five thousand years of civilization history of the Chinese nation bears the great contribution of the Chinese civilization to the world civilization. Chinese civilization and Chinese character culture have a long history, extensive and profound. Chinese classics preserved and handed down by Chinese characters are vast and attract worldwide attention, displaying the cultural characteristics of China's multi-ethnic groups [1], and are the common spiritual wealth of all mankind. English translation of classical works plays an important role in promoting Chinese culture and world cultural exchanges. Under the background of globalization, the English translation of Chinese classics has shown a gratifying development trend, but it still faces many problems in terms of translators' quality, translation level and translation strategies.

As the core content of traditional culture, Chinese classics are not only profound in thought, broad and profound, but also concise and comprehensive, containing rich philosophy, which is an important content of the "Belt and Road" strategic culture. To accurately translate Chinese classics to foreign countries, so that the world can understand Chinese culture, and then truly understand China. The translation of Chinese classics into English is not only an important part of the cultural exchange between China and the West, but also an important way to spread Chinese excellent culture. Chinese classics not only have important ideological value, but also contain rich cultural information. The "Belt and Road" strategy is a strategic concept to revive the Chinese dream, not only a good opportunity for economic rise, but also a historical opportunity for the rejuvenation of China's excellent culture. Under this background, it is of great significance to carry forward the Chinese culture, improve the level of cultural opening and expand foreign cultural exchanges.

Both traditional manual translation methods and Chinese-English translation tools have great limitations and are increasingly unable to meet the translation needs of the information age. Corpora provide a large amount of real corpus for translation and lay an important material foundation for corpus analysis and translation behavior description [2]. In the process of translation, corpora can effectively assist translators to express relatively accurate translations, provide a large number of words,

phrases and even sentence pattern samples, promote the development of translation theories and breakthroughs in translation thinking, and improve translation efficiency. At the same time, corpora can also provide rich language materials in the translation process to help translators improve accuracy and enrich the translation content. The translated works of Chinese cultural classics are valuable wealth and play an important role in reference. The construction of a large-scale parallel corpus with abundant, accurate and reliable data is crucial to a high-quality machine translation system, and makes machine translation move towards the ideal state of "faithfulness", "expressiveness" and "elegance" [3]. The establishment of a corpus for the translation of Chinese classics into English provides a new way for the study of the translation of Chinese classics and the dissemination of Chinese culture.

2. Outstanding Problems to be Solved

Through the sorting and research of relevant literature in the early stage, the following outstanding problems to be solved are summarized.

2.1 The English Translation Level of Chinese Classics is not High

The theoretical construction is not perfect, and the emphasis is still on western translation theories such as "domestication and foreignization", "literal translation and free translation" [4]. In the past, many famous works of Chinese classics translated into English were translated by foreign translators, such as the Analects of Confucius translated by the famous British Sinologist James Legge and the Shuihu translated by the American writer Pearl S. Buck. Most of Mo Yan's works, which won the Nobel Prize in Literature, were translated by Howard Goldblatt, a famous American Sinologist. What foreign translators lack is cultural background knowledge. Lack of English proficiency and inaccurate wording are the biggest challenges faced by Chinese translators. Both foreign translators and Chinese translators lack English proficiency in classical works.

2.2 Corpus Support for Machine Aided Translation

China has a long history and culture, and the classical texts formed in different periods have differences in sentence structure, wording and grammar, which together form the characteristics of The Times of the classical texts and increase the difficulty of machine translation [5]. Usually, the accuracy rate of text machine translation is about 70%, and the accuracy rate of classical translation is even lower, which is far from the real text. As a kind of human-computer interaction system, machine translation can help translators work and improve the efficiency and accuracy of translation. Machine translation requires a corpus to provide support for vocabulary, terms and phrases, and to find the translation of the same or similar sentences from the translated text, avoiding repetitive labour and improving translation efficiency.

2.3 Corpus Included Objects and Internal Structure Problems

The translation of Chinese classics is a large category, and in order to facilitate the study and application, it needs to be classified into several less categories. What kind of internal structure is reasonable and effective, and what kind of criteria is representative of corpus selection, are all problems that must be solved. The translation prototype theory holds that the internal structure of English translation corpus should take professional translation or professional translation as the core of translation category, and both natural translation and non-professional translation can be regarded as the extension of prototype. Different kinds of texts have unequal status, and some emphasis should be placed on corpus collection.

3. Principles of Constructing Bilingual Parallel Corpus for Translation of Chinese Classics into English

The construction of corpus needs to serve specific research purposes. The construction principle emphasizes that the construction of corpora should serve specific research purposes and ensure that the content and structure of the corpus can meet specific research needs. The construction principles provide clear standards and guidance for the construction of the corpus to ensure the high quality and availability of the corpus data. With reference to relevant literature [6,7], the following construction principles are proposed.

3.1 Principle of True Objectivity

The principle of true objectivity means that the text in the corpus should truly reflect the use of the language, fully present the overall characteristics of the target language, and ensure objectivity and representation. In terms of corpus collection, the text comes from the real language environment, and after careful screening and editing, it can effectively reflect the use of language. In terms of the form of corpus, it reflects the practical application of classical translation, mainly written corpus, and in the multimedia context, oral corpus should also be included. In terms of data entry and annotation, the authenticity of the data should be maintained, the original appearance of the data should be faithful to the greatest extent, and the "original copy" should be carried out as far as possible.

3.2 Principle of Dynamic Development

Dynamic development reflects the essential attributes of the world, whether it is the material world or the spiritual world, are constantly changing and developing. Dynamic development emphasizes the internal change and development of things, and also emphasizes the relationship between things and interaction. Translation corpus is also a dynamic development process as translation needs move closer to the target language, which requires continuous in-depth research and exploration to reveal the universal laws of the translation process. It combines cross-sectional corpus collection with longitudinal tracking investigation, pays attention to the integration of multidisciplinary theories, and collects longitudinal diachronic corpus in real time to achieve continuous updating and improvement.

3.3 Principle of Collaborative Innovation

Collaborative innovation requires independent innovation entities to break down barriers, gather information and resources, exchange and interact, and carry out in-depth cooperation and resource integration in order to achieve common goals, thus generating nonlinear effects of system superposition. The translation of Chinese classics is a work involving multiple departments and personnel. The competent departments should provide financial guarantee, and relevant institutions and personnel should realize interactive communication and resource sharing, so that each link can give full play to their respective advantages, jointly carry out bilingual corpus construction, and promote the innovation of translation style research methodology, so as to serve the translation practice of Chinese classics and improve the translation quality.

3.4 Principle of Open Sharing

The open and shared corpus provides researchers with rich linguistic data resources, which is helpful for related personnel to conduct deeper linguistic research and linguistic phenomenon analysis. The bilingual corpus of English translation of Chinese classics has expanded from "independent establishment and self-management" to "cooperative sharing", from the sharing of domestic corpus results to the establishment of a sharing relationship with the corpus of other countries in the world. In addition, through the interactive platform of multilingual corpus, it can promote international cultural exchanges, promote the discovery and summary of theoretical viewpoints, enhance mutual understanding and friendship between different countries, and contribute to promoting world peace and development.

4. An Application Analysis of Bilingual Parallel Corpus in English Translation of Chinese Classics

The research of translation corpus has developed rapidly in recent years, and many translation corpora, parallel corpora and comparable corpora have been built and put into use, and have achieved good results. English translation corpus of Chinese classics can be used not only in translation practice, but also in translation research and translation teaching. The specific application methods in various fields are studied to maximize the application effect of bilingual parallel corpus of Chinese classics translation from English.

4.1 Field of Translation Practice

With the development of information technology, it is easier to use corpus tools to assist translation practice. Compared with traditional paper dictionaries and modern Internet search engines, the unique

advantages of corpus can improve the equivalence between bilinguals, fill the absence of semantic items in bilingual dictionaries, and have the advantage of whole sentence interpretation, which is increasingly helpful to translation practice [8]. Translators can use corpora to acquire professional knowledge, learn expression methods, imitate writing styles, and provide translation strategies for translators in translation practice. The auxiliary effects of corpora on translation practice can be broadly divided into two categories: better understanding of the original text and better generation of the translation. In Chinese-English translation, corpora can also be used to check whether the sentence structure is reasonable. In Chinese-English translation, when Chinese sentences contain multiple identical expressions, English is not accustomed to repetition and prefers diversified expressions, so synonyms are often used. The use of corpus can search for synonyms to help translators choose the right words. The role of corpora in automatic machine translation is mainly reflected in providing rich language data to help machine learning models better understand and generate sentences in the target language. Corpora contain a large number of real language materials, which can help machine translation systems learn the actual use of the language, thereby improving the accuracy and naturalness of translation. In addition, the diversity and scale of the corpus directly affect the performance and effect of the machine translation system, and a rich corpus can improve the generalization ability of the model, so that it can perform well in various scenarios.

4.2 Field of Translation Studies

Corpus translation research is a corpus approach to observe and describe translation, and to explain the essence of translation. Corp-based translation style research can be roughly divided into three categories: from the perspective of corpus translatology, the model construction of translator style research and the choice of research tools. This paper explores the translator's style characteristics through corpus text analysis. Based on the construction of parallel corpora, the translator's style is examined from the three levels of vocabulary, syntax and discourse [9]. Corpus translation research has made remarkable achievements in the exploration of translation universality, and this research method has obvious advantages, transforming translation universality research from an initial intuitive or vague form to a clear or specific and actionable research hypothesis, and from a study only targeting two languages and limited to individual text types to a large-scale systematic comparative study. With the construction and development of corpus, contrastive linguistics studies and translation studies continue to integrate, and the academic community begins to use the composite form of analog corpus and translation corpus to examine the systematic differences between different languages, and use statistical analysis methods to more comprehensively show the characteristics of translators' styles at all levels. Different from the previous studies on the vocabulary and syntax of traditional literary works at different levels through corpus, researchers in recent years have paid more attention to the study of corpus translation in the external translation and publicity of traditional Chinese culture. It mainly adopts the corpus method to compare and analyse the characteristics of translation language, so as to promote the foreign translation of Chinese classics and the external dissemination of Chinese culture, so as to better promote the "going out" of Chinese scientific and technological classics and Chinese culture.

4.3 Field of Translation Teaching

Corpus plays an important role in translation teaching. By introducing corpora, we can provide more practical learning experience, improve students' translation level, enrich teaching content, expand students' vision, and cultivate students' practical ability and intercultural communication ability [10]. Corpora can not only help students better understand the text, but also allow students to apply the knowledge in practice, cultivate language sense and context awareness, and improve the quality of translation. Through reasonable use, it aims to provide students with a more comprehensive learning experience, to promote students' learning and growth. The use of corpus in translation teaching can effectively solve lexical semantic contrast and accurately locate translation problems, so as to help students improve their expression ability and reduce translation errors. At the same time, it is convenient for students to use the auxiliary tool of corpus to quickly and accurately search for specific language expressions, so that the translation of terms and phrases is more accurate, and it can provide multiple translation options or translation references, which is more abundant than the bilingual dictionary. Bilingual translation involves not only vocabulary and rhetorical devices, but also the conversion of various sentence patterns, including existential sentences, non-subject sentences, passive sentences, reduced sentences and elliptical sentences. In formal and written English expressions, sentence patterns usually use the passive voice to make the text appear accurate, formal and

authoritative, while there is no specific requirement for the use of passive voice in written Chinese statements. The application of corpus in translation teaching not only helps to make use of the abundant language materials and words of corpus, but also helps to monitor the learning progress, thus differentiating teaching and guidance, and promoting the sharing of teachers' information, teaching resources and teaching experience.

5. Conclusions

The vast Chinese classics are the crystallization of the ancient Chinese civilization and an important part of the world civilization. People of all countries should share this wealth. However, due to language barriers and ideological differences, the West has long known little about ancient Chinese culture, and the English translation of Chinese classics is an effective means to solve this contradiction. The development of modern information technology provides a new solution for translation practice and translation research. Under the guidance of the basic theories of corpus linguistics, this project comprehensively applies various methods such as literature analysis, expert consultation, statistical testing, interlanguage analysis and linguistic research, and follows the basic principles of corpus construction from the perspective of translators and disseminators, to design a corpus construction scheme for the translation of Chinese classics into English. According to the construction process, the corpus is constructed and the application analysis is carried out. The research results of this topic will help China to build a sound corpus of English translation of Chinese classics, guide the translation practice, translation research and translation teaching of Chinese classics, and serve to promote Chinese culture and promote Chinese classics to the world.

Acknowledgements

This work is supported by Social science planning fund project of Liaoning province in 2021 (No: L21BY002): Construction and application of bilingual parallel corpus for Translation for Chinese classics sub-confucian category.

References

- [1] L. Wei. *Translation, publication and communication of Chinese classics*[J]. *Journal of Liaoning Technical University(Social Science Edition)*, 2024, 26(02): 129-135.
- [2] S. M. Dai. *Research on Intelligent English Translation Based on Corpus*[J]. *Journal of Qiqihar Teachers College*, 2024, 44(05): 154-156.
- [3] L. X. Zhao. *Research on the construction of Chinese-English parallel corpora for machine translation*[J]. *Jiaying Literature*, 2024, 39(07): 78-80.
- [4] L. J. Zhang. *Research on the training of Chinese classics translation talents*[J]. *Journal of Anyang Institute of Technology*, 2023, 22(05): 45-48.
- [5] M. C. Wu, L. T. Lin, D. Hu, et al. *A Study on Machine Translation of Ancient Chinese Books from the Perspective of Temporal Characteristics*[J]. *Library Tribune*, 2024, 44(10): 93-102.
- [6] Z. L. Zhou, D. S. Ren. *A Corpus based Study on the Translation and Communication of Ocean Discourse with Chinese Characteristics: Framework, Contents and Principles*[J]. *Journal of Ocean University of China(Social Sciences)*, 2022, 25(04): 23-31.
- [7] Z. Y. Xu. *The Korean Inter-language Corpus Construction plan of Chinese Learner*[J]. *Journal of Kunming University*, 2018, 40(01): 127-132.
- [8] J. L. Liu, J. P. Huang, Y. T. Cai, et al. *Advantages of IT parallel corpus-assisted translation practice*[J]. *English Square*, 2014, 5(11): 34-35.
- [9] F. Qin. *Hot spots and trends in domestic corpus translation research (2013-2023) - multi-dimensional visual analysis based on CiteSpace*[J]. *Overseas English*, 2024, 25(18): 80-83.
- [10] W. F. Wang. *Research on Foreign Language Translation Teaching Mode in Colleges and Universities Based on Corpus Translation Technology*[J]. *Journal of Jinzhou Medical University (Social Science Edition)*, 2024, 22(04): 101-104.