

Construction And Application Of Chinese-English Parallel Corpus Of Jiangxi Provincial Museum Texts Under The Belt And Road Initiative

Chunlian Liu

School of Foreign Languages, East China Jiaotong University, Nanchang, Jiangxi, China

Keywords: Parallel Corpus; Jiangxi Province; Museum Texts; The Belt And Road Initiative

Abstract: Museums are one of the important carriers for inheriting the historical and cultural heritage of mankind, and the translation quality of their commentaries will directly affect the perception of local history and culture by international tourists. Based on the current situation of the English translation of commentary materials in Jiangxi Provincial Museum, this article focuses on the design and construction of the English-Chinese parallel corpus of Jiangxi Provincial Museum texts, and further points out the application of the corpus and the areas for improvement, which will improve the translation quality of Jiangxi Provincial Museum's commentary. It is of great significance to enhance Jiangxi's international image.

1. Introduction

The Belt and Road Initiative is a development strategy of the CPC Central Committee and the State Council to actively respond to the profound changes in the global situation. Jiangxi is an important commodity export destination of the ancient Silk Road. The development strategy of the Belt and Road Initiative has brought great historical opportunities to Jiangxi's cultural "going global". The cultural relics displayed in the museum are important witnesses to the long history of a place and represent the cultural characteristics of a place. Understanding the culture of a region starts from the museum. The history, culture, folklore and other factors it carries are unavailable or even unfamiliar to other places and regions. It is the presentation of the original culture. The text of the Jiangxi Provincial Museum is an important medium for the outside world to understand the history and culture of Jiangxi, and conveys the characteristic culture of Jiangxi region. In view of this research status, the research team intends to use the research as an opportunity to build a Chinese-English bilingual corpus of Jiangxi Provincial Museum. On this basis, this paper tries to provide some guidance and reference for the standardization and accuracy of English translation of related texts and explore the language characteristics of the Chinese-English museum text; so as to establish a good international image of Jiangxi province, attract foreign tourists, and promote the rapid development of Jiangxi's tourism industry.

2. Construction of Chinese-English Parallel Corpus of Jiangxi Provincial Museum

A corpus refers to an electronic database composed of naturally occurring corpus collected according to a certain sampling method. It is a collection of language application materials selected and arranged in an orderly manner according to the research purpose and corpus selection method^[1]. The language in the corpus is the real language material, which provides important resources for language research. At present, the commonly used corpora in translation studies mainly include target text corpus, analog corpus and parallel corpus^[2]. Existing translation corpora are mostly literary corpora or large-scale general corpus mainly containing encyclopedia texts, while museum-type stylistic translation corpora or bilingual corpus translation study corpora are not common. No matter at home or abroad, relatively large-scale bilingual specialized corpora are still rare. Parallel corpus refers to two kinds of corpora stored in the same corpus, and they are in correspondence. One text is the translation of the other text. The construction process of the Chinese-English bilingual parallel corpus of Jiangxi Provincial Museum text consists of three steps, including corpus selection and entry, corpus segmentation and labeling, and corpus input and alignment. The specific

operations are as follows.

2.1. Corpus Selection and Entry

As is known to all, the first issue that needs to be considered in the construction of any corpus is the representativeness of the corpus construction, which is the most basic issue of corpus selection [3]. Therefore, in the construction of this corpus, the author mainly adopts two ways of corpus source: practical investigation and literature investigation. To be specific, the author first took a field trip to the Jiangxi Provincial Museum and captured a large amount of real first-hand corpus. After collecting and collating the corpus with OCR software, the author found that the quality of the English translation of the Jiangxi Provincial Museum's commentary text is uneven, and there are three types of translation errors, such as grammatical errors, spelling errors, and Chinglish. In order to make the source of the corpus more standardized and comprehensive, the research team analyzed, revised and retranslated the English translation of the Jiangxi Provincial Museum's commentary text with reference to the English translations officially released by CNKI, similar museums and existing related corpora. The relevant text commentary in Chinese and its corresponding correct English translation are saved as Word text. Subsequently, the Emeditor and Editplus software were used to delete some "noise" information that may exist in the text, such as extra spaces and blank lines, for subsequent input into the corpus.

The original corpus comes from the pictures taken on site in the museum, which requires scanning or manual input. All the input corpus, whether it is in ".doc" format, ".pdf format", or other formats, are converted to ".txt format for archiving. The text belongs to the corpus corresponding to the translation relationship, and the same file name with adding different suffixes to indicate the parallel versions. At the same time, all file names are named with the Chinese pinyin abbreviation. For example, a Chinese corpus containing "Jiangxi province situation" can be named "JXSQ_CH", and the corresponding English corpus is named "JXSQ_EN". When all corpora are named and archived, records are made, and corresponding files are established, and the source and content of the corpora are briefly explained for future search and update.

2.2. Corpus Word Segmentation and Labeling

In order to search Chinese corpus in the corpus software, you need to segment the Chinese text or add spaces between each Chinese character. The easiest way is to add spaces to the Chinese characters through the "Find" and "Replace" functions of MS word. For research needs, all corpora can be annotated. The annotation/tagging in this study is divided into two categories: heading and in-text annotation/tagging and part-of-speech annotation/tagging. After the corpus is converted to plain text and archived, a lot of information (such as the layout, format, pictures, etc.) in the original corpus of the original text will be removed. Through the title and in-text annotations, the original information can be reflected in the processed documents. Referring to the annotation method of the general Chinese-English corresponding corpus [4], and combining with the purpose and characteristics of this research, we set the title content to include : Title (Chinese and English), author's name, translator's name, style, time, publication information, word count, etc.. Intra-text annotations mainly include subtitles, paragraph boundaries, sentence boundaries, sentence labels and so on. Header and text tags are marked in XML format, as shown in the following table 1:

Table 1. Examples of Headings and In-text Annotations

Content	Mark Symbol	Content	Mark Symbol
Chinese Title	< TITLE_CH >... < / TITLE_CH >	Paragraph Mark	< P >... < /P >
English Title	< TITLE_EN >... < /TITLE_EN >	Sentence Mark	< S >... < /S >
Subtitle	< SUBTITLE >... < /SUBTITLE >	Source	< SOURCE >... < /SOURCE >
Author Name	< AUTHOR >... <	Time	< TIME >... < /TIME

	/AUTHOR >		>
Translator's Name	< TRANSLATOR >... < /TRANSLATOR >	Word Count	< WORDS > ... < /WORDS >
Style	< STYLE >... < /STYLE >	...	

In order to facilitate further retrieval and research in the future, the research team have separately marked part-of-speech tagging on the Chinese and English corpora. For English corpus, the Tree Tagger tagging system is adopted. For Chinese corpus, the ICTCLAS system of the Institute of Computing Technology of the Chinese Academy of Sciences and the YACSI 0.9 Software^[5] for part-of-speech tagging. After the Chinese and English corpus is annotated by the above software, a lot of careful manual proofreading is needed to improve the accuracy of the annotation.

2.3. Corpus Input and Alignment

After the corpus is input and cleaned up, it is necessary to enter the key step of parallel corpus construction, corpus alignment. Although many scholars, such as Huang Junhong etc.^[6] believe that the smaller the alignment unit, the more language information will be provided, and the greater the value of the application will be. However, some experts^[7] pointed out that “Because Chinese and English belong to different language families, there are obvious differences in word order, sentence structure, and logical and semantic expression, etc.. It is impossible to correspond to words one by one in the whole corpus. If it is done, the corresponding difficulty at the sentence level is also quite large.” Therefore, in the construction of this corpus, the research team designed to achieve sentence-level alignment, supplemented by paragraph alignment. There are many software that can realize automatic alignment of Chinese and English sentences, such as WinAlign in TRADOS Translator's Workbench, Alignment tool in ParaConc software, MMAX^[8] program, Abby Aligner and some other self-developed software. But most of the software automatic alignment accuracy is not high, and the operation is more complicated, and a lot of proofreading and manual operations are still required after alignment. Therefore, in the actual operation process, Editplus text editing software and the related functions of MS word and Excel are mainly used to perform sentence alignment operations, and repeat proofreading after the sentence is aligned to ensure the alignment accuracy. After the Chinese and English corpora with sentence alignment are saved separately, ParaConc, the parallel corpus retrieval software, can be used for retrieval and processing.

3. The Application Value of the Chinese-English Parallel Corpus of Jiangxi Provincial Museum

The multi-dimensional study of museum texts is ultimately for the inheritance and dissemination of culture. The construction of a Chinese-English parallel corpus of Jiangxi Museum text materials can improve the quality and efficiency of the museum's propaganda. The language materials in the corpus can be used not only for language research, but also for Jiangxi tourism promotion and other fields, making the corpus play a greater role of resources as follows:

3.1. Cultural Propaganda

Self-built Jiangxi Provincial Museum text corpus study and solve the existing problems of English translation of museum texts, and provide materials and references for future publicity of historical and cultural heritage. This can create an efficient and standardized English-Chinese parallel corpus of foreign publicity texts of historical and cultural heritage in Jiangxi under the new situation of the Belt and Road Initiative, and build a parallel corpus of English translations of various museum texts in Jiangxi Province as well as a Chinese-English translation corpus of Jiangxi local cultural libraries. The construction lays the foundation and provides a reference for the translation research of the foreign propaganda materials of the Haihunhou Kingdom Relics Park

which is under construction. The Chinese-English parallel corpus of Jiangxi Museum's text is conducive to the inheritance and dissemination of Jiangxi's history and culture, making the splendid ancient culture of Jiangxi well known to the world, enhancing Jiangxi's international image, serving the local economy, and promoting the development of Jiangxi's tourism economy .

3.2. Translation Studies

The self-built Jiangxi Provincial Museum's textual English translation corpus can objectively describe the translation phenomenon or translation facts in this field; and achieve summary and induction on the basis of a large number of real corpus observations and data analysis. The organic combination of quantitative and qualitative research of museum interpretation texts, as well as bottom-up and top-down methods can be realized through the introduction of quantitative research. At the same time, the corpus can provide materials for the study of cultural heritage translation, so as to lay the foundation for the future theoretical research in the translation of historical and cultural heritage, such as the study on the commonality of translation, the characteristics of translation language, the study of translators' style and the study of translation norm. From this, it will be extended to the study of the translation of museum-type publicity materials and the study of Jiangxi regional culture, which has positive academic significance.

3.3. Translation Teaching

The Chinese-English bilingual parallel corpus can also be used in the teaching of Chinese-English translation for English majors or ESP English teaching for liberal arts majors. The real, lively, intuitive and rich teaching materials provided by the corpus can promote learners' understanding of Chinese-English travel texts from the micro level (characters, words) to the macro level (sentence connections, discourse structure, and rhetorical features), encourage them to read and discuss the similarities and differences in language features between Chinese and English parallel corpora and summarize practical translation strategies to apply to classroom teaching, which can stimulate students' enthusiasm for learning and truly realize data-driven learning. In ESP English or translation teaching in local colleges and universities, the application of these teaching materials will undoubtedly help improve learners' understanding of local history, geography, and culture, and also help enhance learners' cross-cultural communication skills.

4. Conclusion

In the context of the current era, "serving the national strategy and serving the Chinese culture going global" is the historical mission of language translation and its industry. Jiangxi has a lot to do to build cooperation regional value bond based on cultural memory. Constructing a Chinese-English parallel corpus of the Jiangxi Provincial Museum's commentary texts and improving the language services of the Jiangxi Provincial Museum, can greatly enhance translation efficiency and ensure translation quality, expand the propaganda of Jiangxi's regional history and culture, and vigorously promote the "Jiangxi Scenic Unique" tourism brand. Tell the story of Jiangxi well, enhance the international reputation of Jiangxi tourism, so that the culture of Ganpo (Jiangxi) can go to the world better, and promote the spread of Jiangxi's history and culture and the exchange of culture with the world.

Acknowledgements

Jiangxi Provincial Social Science Planning Project 2018 General Project "Research on the Construction of the English Translation Corpus of Jiangxi Provincial Museum Texts under the Background of "The Belt and Road" Initiative (18YY09)

References

[1] Hu Kaibao. Introduction to Corpus Translation [M]. Shanghai: Shanghai Jiaotong University

Press, 2011.

[2] Lu Hefa, Jiang Lu, Zhou Jianbo, etc. Research and Construction of Public Sign Translation Corpus [J]. Contemporary Foreign Language Research, 2015(10): 48-55.

[3] Biber, D. Representativeness in corpus design[J]. Literary and Linguistic Computing, 1993, 8 (4): 16-17.

[4] Wang Kefei. Design and construction of a new bilingual corresponding corpus[J]. Chinese translation. 2004(6):73 -75.

[5]Wu Liangping. Yacsi 0.9: Yet Another Chinese Segmenter Inter-face to ICTCLAS. 2012. Changsha: Hunan University of Commerce.

[6] Huang Junhong, Fan Yun, Huang Ping. A Review of Bilingual Parallel Corpus Alignment Technology [J]. Audio-visual Foreign Language Teaching, 2007(6): 21-25.

[7] Wang Kefei. Design and construction of a new bilingual corresponding corpus [J] . Chinese translation. 2004(6):73 -75.

[8]Liu Zequan, Tian Lu, Liu Chaopeng. Creation of Chinese and English Parallel Corpus of "A Dream of Red Mansions"[J]. Contemporary Linguistics. 2008(4):329 -339.