

Corpus-based Intangible Cultural Heritage Translation Path and Strategy Analysis of Chinese Culture "Going Out"

Yanjun Zhou

Beihai Campus, Guilin University of Electronic Technology

Abstract. Intangible cultural heritage translation is regarded as a new subject in the new era of economic construction development, can not only the external publicity of the rich national culture, but also improve our country's cultural propaganda comprehensive level. By studying the corpus-based intangible cultural heritage translation path, the quality and efficiency of cultural translation can be comprehensively improved, and the reliable translation principles and effective strategies of large-scale data can be defined, which can provide effective basis for international cultural publicity in the new era. Therefore, on the basis of understanding the research status of corpus and intangible cultural heritage translation, this paper mainly discusses the corpus-based machine translation system, and then analyzes effective strategies for Chinese culture to "go out" from the perspective of practical development.

Keywords: Corpus; Intangible cultural heritage translation; Chinese culture; Machine translation; statistics.

1. Introduction

In the construction and development of modern society, intangible cultural heritage belongs to the basic component of local cultural construction and communication, bringing together a number of basic cultural elements such as local ethnic customs, traditional culture and technological means. Strengthening the publicity and promotion of intangible cultural heritage can not only improve the local social awareness, but also contribute to the rapid development of local market economy. Nowadays, the pace of internationalization of Chinese cities is getting faster and faster. The external publicity of intangible cultural heritage is very important for the construction of local cultural projects. As this work involves text translation, how to improve the quality of text translation is the core of current scholars' research. By using scientific and technological methods to study the translation of non-legacy texts, we can comprehensively understand the effective strategy of "going out" of Chinese culture, which can provide effective basis for economic construction and urban innovation development in the new era. Although text propaganda of intangible cultural heritage is an important means of national culture communication, from the perspective of practical development, there are not many researches on this subject at home and abroad, and the number of relevant papers is very small. For example, Chen Fangrong mentioned the translation of intangible cultural heritage texts for the first time in 2011. [1-3]Huang Dongqun proposed in his follow-up studies that the text translation of intangible cultural heritage should be based on the characteristics of cultural translation, and a variety of translation means should be used reasonably. After mastering important cultural information, the differences between Chinese and Western cultures should be accurately determined, so that more audiences can understand the rich and colorful Chinese culture. Yang Yonggang proposed in his research that intangible cultural heritage has profound cultural deposits and unique textual features. Therefore, based on the investigation of Yangzhou intangible cultural heritage and the current situation of Chinese-English cultural translation, based on the functional translation theory, this paper mainly discusses how to apply special English translation skills to translate non-relic texts, so as to achieve the basic goal of translation specialization. From the perspective of overall development, Chinese study on non-legacy text mainly focuses on intercultural translation methods and techniques. The universal translation principles and application strategies cannot be summarized precisely because the proposed research topics are limited and the theoretical knowledge and research achievements are not much.[4-6]

In the 1980s and 1990s, researchers in the Netherlands and Sweden conducted research on the characteristics of translated languages based on the analogical corpus of original novels and translated novels, which was also the earliest exploration direction in the field of corpus and intangible cultural heritage translation path research. After that, some scholars put forward relevant theories of corpus linguistics and translation studies, focusing on the specific methods and development prospects of applying corpus in translation studies, which is also regarded as a landmark event in the establishment of a brand new research model by the academic circle. In the late 20th century, some scholars established corpora corresponding to the original text and the target text, and selected new technical means to show the essential characteristics of the target text. This new model effectively broke through the traditional orientation of fragmented text research in translation studies, and began to choose large-scale corpus data with computer technology as the basis. This provides a new idea for the study of text translation in the new era. Generally speaking, applied corpora in translation studies can be divided into two main fields, one is theoretical research, the other is practical research. This paper, after understanding the machine translation system of intangible cultural heritage based on corpus, mainly explores the translation path of intangible cultural heritage based on corpus, according to the basic requirements of cultural communication and development in the new era, so as to provide effective basis for cultural communication and economic construction.

2. Methods

2.1 System Analysis

In the late 1940s, Weaver proposed in his memo to complete machine translation by means of decoding passwords. This method is essentially a statistical method, which mainly uses statistical methods to deal with the problems existing in machine translation. Since there were no high-performance computers and connected corpora at that time, machine translation technology with statistics as the core was not mature. With the rapid development of modern science and technology, both computer technology platform and connected corpus have been comprehensively optimized, so some scholars began to continue to study machine translation systems based on statistics in the 1990s. The machine translation problem of this method is regarded as a noisy channel problem, and the actual model is shown in Figure 1 below:[8-9]

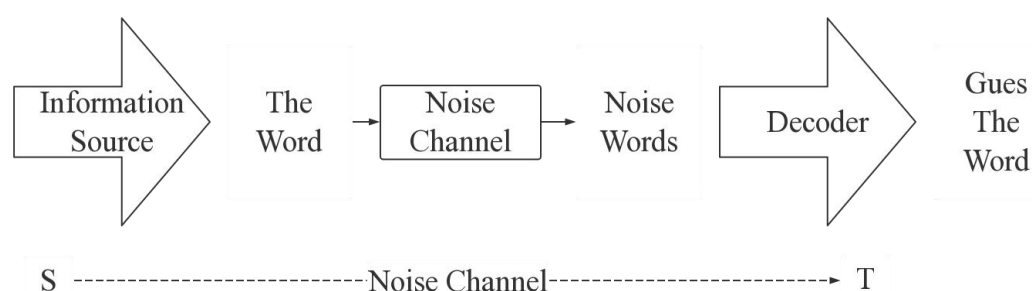


Figure 1. Model diagram of noisy channel

Based on the analysis of the figure above, it can be seen that a language S distorts and changes after the noise of one channel, and will appear as another language T on the other side of the channel. Thus, the translation problem is how to recover the most likely language S according to the observed language T. In practice, researchers should focus on solving ideological problems when studying statistical machine translation systems: first, to estimate the probability of language model $P(T)$; Secondly, the translation probability $P(S/T)$ is estimated. Finally, an effective search algorithm is designed to solve T and ensure that $P(T)P(T/S)$ is maximized. Taking IBM's Candide system as an example, researchers developed and designed an English-French machine translation system based on the bilingual corpus of Candide and Canadian parliamentary debate records. The

English-French machine translation system was developed and designed. The specific structure is shown in Figure 2 below:

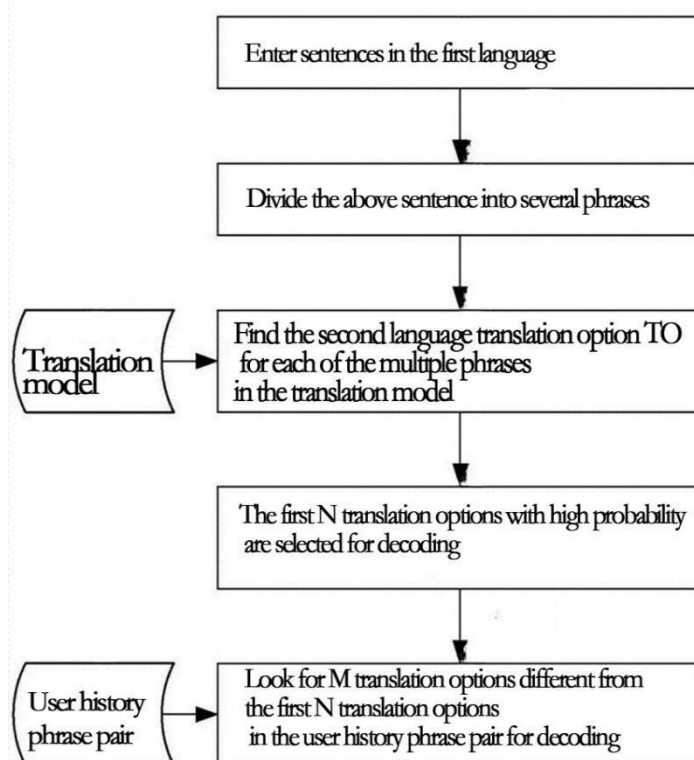


Figure 2. Structural diagram of the English-French machine translation system

Based on the analysis of the above table, it can be found that the overall system design includes three parts: the first is the ternary grammar model for English, the second is the ternary grammar model for French, and the last is the high-quality correspondence model for English and French partially aligned sentences. Although from the perspective of practical application, this method can get the result of machine translation, but the sentence accuracy is only 40%, which cannot solve all the difficult problems.[10-12]

2.2 System Improvement

On the basis of studying statistical machine translation system, some scholars have proposed a system based on maximum entropy method. The advantage of this application method is that only the appropriate feature set needs to be selected during the modeling period, and the model should reflect the contribution of these features to the model in the form of parameters. The feature selection will not be restricted by historical conditions, and can be flexibly applied during the research operation. At the same time, when dealing with problems such as excessively sparse data, the maximum entropy model does not need to use a special smoothing algorithm, but chooses to realize data smoothing according to features. In a machine translation system based on maximum entropy method, the translation model in the noisy channel model should be transformed into a reverse translation model first, so as to comprehensively simplify the search algorithm and ensure that the application performance of the final system will not be degraded. Taking the translation system of Language Weaver as an example, it mainly follows the process as shown in Figure 3 below.

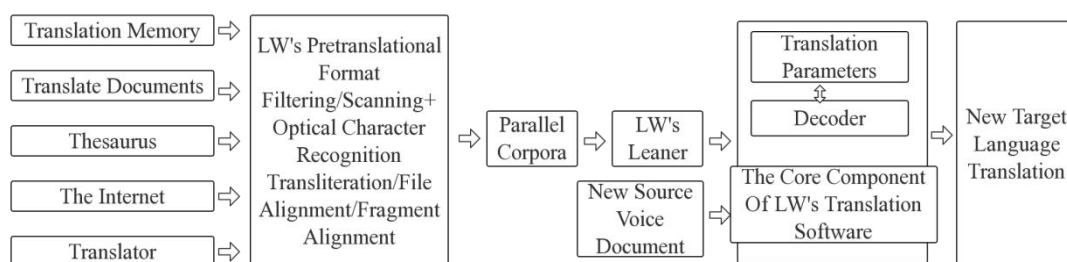


Figure 3. Flow chart of Language Weaver's translation system

Now, this system has become an important basis for machine translation research. After mastering its core content, different types of translation machines are proposed. At the same time, Google enterprise also put forward a new statistical machine translation system, the actual translation effect is better than the original machine translation system, the overall design is divided into two parts, on the one hand refers to training, on the other hand refers to decoding, the actual training process is shown in Figure 4 below:

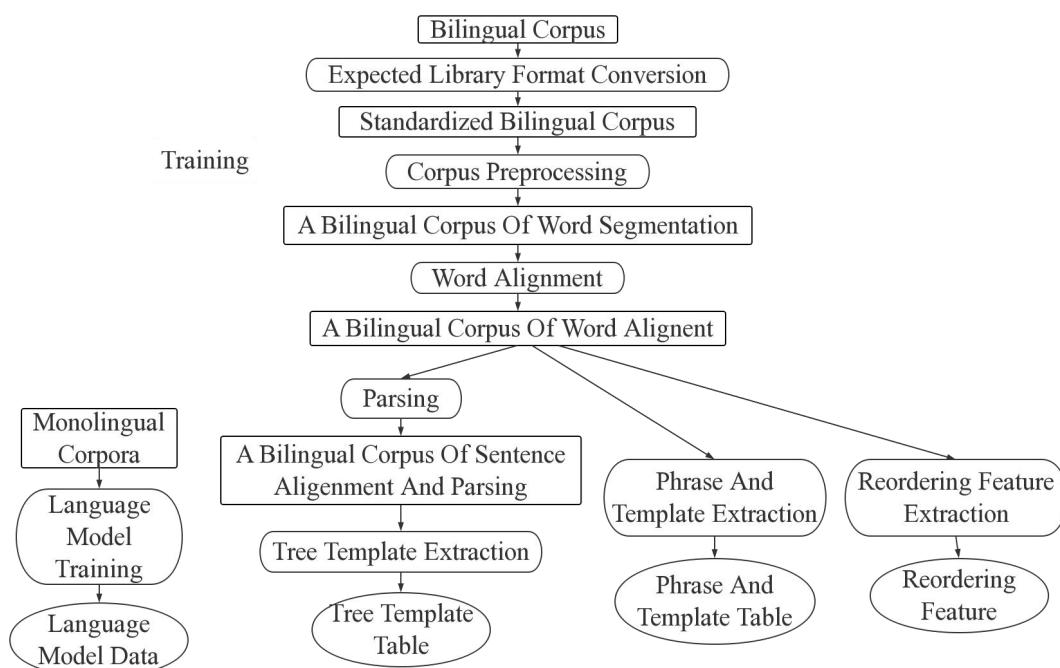


Figure 4. Flow chart of machine translation training

The actual decoding process is shown in Figure 5 below:

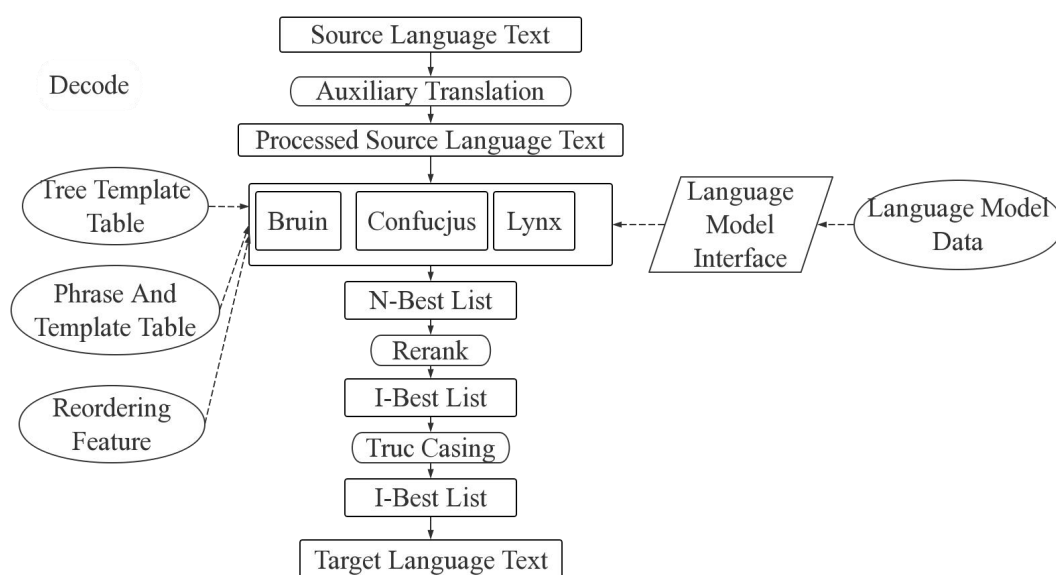


Figure 5. Decoding flowchart of machine translation

According to the above figure, language model data is obtained from the model, while phrases, template tables, and reordering features are obtained from model translation. During machine translation decoding, n-best was obtained, followed by a 1-best list Rerank, and True-casing the target language text. Nowadays, statistical machine translation research at home and abroad has achieved excellent results, and occupies an important position in the development of Chinese cultural communication.

3. Result analysis

According to the structural analysis of the machine translation system of intangible cultural heritage statistics studied in this paper, taking Chinese-English translation as an example, the current corpus-based intangible cultural heritage translation path and Chinese culture "going out" strategy are mainly reflected in the following points:

Firstly, three corpora are constructed. In order to effectively study the main contents of translation of Chinese non-posthumous texts into English, it is usually necessary to establish at least three special purpose corpora. The first is the Chinese-English parallel corpus of non-posthumous texts, the second is the English-Chinese comparable corpus of non-posthumous texts, and the third is the English comparable corpus of non-posthumous texts. A large number of original English non-posthumous texts, original Chinese non-posthumous texts and English texts translated from Chinese non-posthumous texts should be collected in these corpora.

Secondly, determine the research content. In the corpus-based intangible cultural heritage translation path, scientific statistical methods should be used to compare and analyze the language features and discourse features of English and Chinese non-legacy texts, and actively seek the laws of the two texts in language, text and discourse creation, so as to provide scientific methods and theories for the translation of flying fish texts. For example, in the comparative study of language features, according to the analysis of the comparable corpus selected in this paper, the language features of the two types of texts can be compared by using data and other intuitive ways, so as to clearly present the same or different language styles of the two types of texts, so as to provide the basis for the successful language conversion.

Finally, master the research methods. In corpus translatology text research, the most common method is the relevant content proposed based on data statistics, which is also the main way to apply corpus to the translation of non-posthumous texts. Before the research, a sizable corpus is established, and then the conversion rules of the two languages are found in the parallel comparison between the original text and the target text. At the same time, a non-posthumous English-Chinese

comparable corpus and a non-posthumous English comparable corpus are also built. Finally, after mastering the basic language model of non-posthumous translation, the basic principles and technical methods of translation are summarized and analyzed. From the perspective of practical application, quantitative research, as a traditional positivism scientific research method, is to carry out statistical analysis on individual variables and conditions under the general situation, and finally explain relevant things and specific phenomena. Due to the scientific and perfect data information of statistical analysis, it is feasible to combine quantitative and qualitative research on the translation of non-legacy texts, which provides a reference for the study of translation techniques and methods in the new era.

4. Conclusion

To sum up, the translation of non-legacy texts is an important part of the transformation of China's economic construction and cultural transmission into the modernization process, as well as the main channel for the mutual integration of Chinese culture and world civilization. As intangible cultural heritage is a fundamental part of the development of a country and a nation, the study of machine translation using corpus is rather complicated. It is necessary to obtain practical translation methods and technical principles after mastering scientific and perfect data statistics and analysis, so as to achieve the ultimate goal of cultural transmission while guaranteeing the quality of text translation. Therefore, in the study of non-posituaries, many projects are put forward according to the basic principles of Chinese culture "stepping out", which not only strengthens the innovation level of the machine translation system, but also masters more translation rules in the construction of translation discipline.

5. Acknowledgement

*Project: This paper is one of the phase results of the research project "Research on the Influencing Mechanism and Improvement Measures of the Translation Effect of Intangible Cultural Heritage in Guangxi (22FYY014)".

References

- [1] Yao Yan, Minkang Zhou. An analysis of the Spanish translation of TCM Terms based on the International Standard of Chinese and Western Comparison of Basic Terms of TCM [J]. Chinese Science and Technology Terms, 2021, 23(3):7.
- [2] Fanxu Meng. Feasibility analysis and Mode Exploration of Integrating Local Intangible Cultural Heritage resources into the Talent training of Master of Translation and Translation (MTI) [J]. Journal of Liaoning University of Science and Technology, 2021(2):33-35.
- [3] Zhongliang Zhou, Dongsheng Ren. Corpus-based Research on Marine Discourse Translation with Chinese Characteristics: Framework, Content and Principles [J]. Journal of Ocean University of China: Social Science Edition, 2022(4):9.
- [4] Kai Ding and Jun Wang. A comparative study on the English translation of Li Yu's Ci based on Corpus [J]. Journal of Heilongjiang Institute of Technology, 2022, 36(5):52-57.
- [5] Jia Wang, Jing Xiong. On the construction path of the Chinese-English parallel Corpus of Folklore on Jiangxi Provincial Intangible Cultural Heritage List [J]. Modern Ancient Culture and Innovation, 2021(29):2.
- [6] Qiunan Zhang. An Exploration on the Formation Mechanism of English Expressions of Chinese Characteristic Words Based on China-related English Corpus [J]. English Square: Academic Research, 2022(33):4.
- [7] Mingwu Xu, Pei Wang. Corp-based research on the English translation of Chinese scientific and technological classics: Current situation, significance and prospect [J]. Foreign Languages and Foreign Language Teaching, 2022(5):10.

- [8] Ying QinJieyu Qin. Research on the English translation strategies of dialect words in Tumen based on Corpus [J]. Times People, 2021, 000(006):P.1-2.
- [9] Lan Chen,Lin Tan,Chunhua Chen, et al. Planning and development Path Based on Spatial structure characteristics of Intangible Cultural heritage: A case study of Mianyang City [J]. China Chengcheng, 2023, 37(3):7.
- [10] Yutao Yang,Ran Yang. Construction and application of a small multimodal parallel corpus of Martial Arts Culture in Central Plains [J]. Sports Vision, 2022(20):3.
- [11] Yiping Wen,Liping Jiang. Application of Human-computer Interaction in the English Translation of Intangible Cultural Heritage Texts: Based on Bilingual Corpus Research. Language and Cultural Studies, 2022(2):5.
- [12] Liping Jiang. A study on the Linguistic features of English translation from the national Intangible Cultural Heritage parallel Corpus [J]. English Abroad, 2022(24):3.
- [13] The Army of Qin. Characteristics and thoughts on Foreign translation of Guangxi Intangible Cultural Heritage [J]. Journal of Kaifeng Vocational College of Culture and Art, 2021, 41(11):35-37.
- [14] Tao Fan,Hao Wang,Baolong Zhang. Research on Intangible Cultural Heritage Text Attribute Extraction Based on Remote supervision and Deep Learning [J]. Information Theory and Practice, 2021, 44(10):8.
- [15] Ping Duan. Corporate-based Translation of Corporate Social Responsibility Reports of Chinese Listed Companies [J]. Modern English, 2021(23):4.