

Original Paper

Research on the Construction of a Chinese-English Parallel Corpus for Terminology of China's Intangible Cultural Heritage

Shi Shi & Yiduo Wang

School of Foreign Languages, Hubei University of Technology, Wuhan, Hubei, China

Received: January 26, 2026

Accepted: March 16, 2026

Online Published: April 1, 2026

doi:10.22158/selt.v14n2p1

URL: <http://dx.doi.org/10.22158/selt.v14n2p1>

Abstract

Intangible Cultural Heritage (ICH) is regarded as an important carrier of the cultural genes of the Chinese. Whether the translation of terms is standardized will directly affect the effectiveness of international communication. At present, there are still many problems in this field. Cross-cultural communication and translation research lags behind. Chinese-English parallel corpora are also very scarce. This can no longer meet the needs of international dissemination and cross-cultural communication of intangible cultural heritage. This study focuses on the construction of a Chinese-English comparative corpus of intangible cultural heritage terms. This section describes the corpus construction process in detail, including planning and design, text collection, data preprocessing, bilingual alignment, and storage management. The purpose of this study is to build a standardized and practical parallel corpus and promote better cross-cultural dissemination of China's intangible cultural heritage.

Keywords

ICH, Chinese-English parallel corpus, corpus construction, cross-cultural communication

1. Background to Constructing the Chinese-English Bilingual Corpus of ICH Terminology

1.1 Core spirit of China Intangible Cultural Heritage

Cultural heritage is invisible and reflects the cultural memory of countries, and it is an important repository of the historical and cultural success of a country (Zhang Li, 2025). It consists of 10 main categories, such as traditional crafts and folk traditions, traditional theatre, traditional music, folk literature, and traditional medicine. It includes 43 world-level items of intangible cultural heritage and 3,610 national-level items of intangible cultural heritage. This forms a comprehensive system. Its main spirit is not only the skill or custom, but rather its values. Forming thinking, aesthetic, and cultural memory. It is the spiritual glue that maintains the constant evolution of the Chinese culture. In addition,

it is one of the irreplaceable cultural and historical values that differentiate it from other cultural systems.

1.2 Background to the International Dissemination of ICH

Transmission of intangible cultural heritage to international frontiers is an important platform through which Chinese culture is promoted. It contributes to the soft power of the country and mutual learning of Chinese and other cultures. It is such that the world can learn more about the Chinese past and national spirit, dismantling cultural boundaries. At the same time, it provides an impetus to the living cultural heritage and enhancing confidence in national culture. Due to the profound development of the Belt and Road Initiative and the International Dissemination Project of Chinese Culture, the tendency in policy favoring translation of intangible cultural heritage to foreign markets and the needs of the world in such products will grow.

The necessity to distribute information regarding intangible cultural heritage is becoming more and more acute in the framework of international cultural interactions, cultural tourism collaboration, and academic conferences. Technical, or terminology, which is the most fundamental unit of expressive content in intangible cultural heritage, plays a key role in capturing the essence of the intangible and ensuring its expressive accuracy. This is because standardization provides direct influence on the quality of dissemination of Chinese-English translation. The existing deficit of standardized Chinese-English terminology aids does not suit the practical needs of ICH spreading on the international level, which can be considered the impulse to create a Chinese-English ICH terminology corpus.

2. Developments in Chinese-English Translation of ICH Terminology and Corpus Building Report

2.1 Cross-Cultural Communication and Translation of ICH Lagging

Recent domestic ICH studies are mainly concentrated on the aspect of preservation, the passing of skills and cultural meaning. There is a strong lag in systematic studies on ICH cross-cultural communication and Chinese-English translation. On the one hand, studies on ICH translation through a cross-cultural communication perspective are shallow.

The majority of literature is narrowed down to case studies, with none of them incorporating the concepts of cross-cultural communication and the cognitive habits of the reader. This leads to a phenomenon that translation should combine with cultural fidelity and communicative adaptation. Nevertheless, research on ICH translation lacks of practice. Researchers are lack of specific backgrounds or experience in cross-cultural communication in the field of Non-cultural, whereas translators are lack of theoretical instructions. At the same time, the standard of translation has not formed. All of these lead to a number of problems, including confusing translation of terminology and bias in cultural transmission. Difficult to satisfy the necessity of spreading ICH in communication.

2.2 Underdevelopment of the Corpora in ICH

There is an increasing interest in corpus. With the development of computer and AI, the study on corpus has been flourishing (Zhou Ying, 2018). However, in the context of the acceleration of the corpus process, the current situation of corpus construction related to intangible cultural heritage is out of touch with the actual needs. They as a whole present the disadvantages of scarcity, homogenization, and single function.

At present, the number of corpora related to intangible cultural heritage that has been established in China is very small, and most of the corpora do not focus on professional terminology, and most of the related translation work is based on complete text translation. Intangible cultural heritage terms have not been stripped, sorted out and standardized from the text.

As a result, the core need for retrieving terminology and obtaining standardized translations in translation practice and academic research cannot be met. Moreover, many of the terms of intangible cultural heritage with local characteristics have not been covered. In addition, some translations have not been double reviewed by intangible cultural heritage experts and translation experts, and their authority and standardization are not enough. Therefore, it is necessary and urgent to build a special Chinese-English comparison corpus of intangible cultural heritage terms.

3. Construction of Chinese-English Comparison Corpus of ICH Terms

3.1 Planning and Design

First of all, the coverage and type of corpus should be clarified. In view of the current narrow corpus coverage, we need to delineate the coverage of corpus according to the ten categories of intangible cultural heritage. It should include intangible cultural heritage classic documents, technical processes, folk activities and other content in core areas such as traditional skills, folk customs, and traditional dramas.

At the same time, both well-known terms and local characteristic terms are required, so as to solve the problem of uneven corpus coverage. The next step is to design a scientific database. For issues with single-function databases, design a specialized database to store Chinese terminology, English translations, annotation information, alignment relationships, etc., which facilitates everyone's search for terminology and standardized management of terminology.

Additionally, establish uniform quality standards. Regarding the problems of uneven quality and lack of authority, the translation quality must be strictly standardized. The final translation must be accurate, professional, and suitable for cross-cultural communication. For words with cultural load, corresponding cultural annotations should also be added. Through a series of operations, the subsequent corpus collection, processing and review will have clear basis. The authority and standardization of the corpus can also be ensured, avoiding the core shortcomings of corpus creation from the very beginning.

3.2 Data Collection

Chinese data mainly comes from libraries, archives, intangible cultural heritage inheritance bases, etc.,

including classic works on ICH and memoirs of ICH skills. The English data mainly comes from international media reports, bilingual texts of the UNESCO Intangible Cultural Heritage List, and intangible cultural heritage translations reviewed by experts. The sources of corpus are strictly limited to four categories, namely the UNESCO Intangible Cultural Heritage List, bilingual documents of the China Intangible Cultural Heritage Protection Center, research results of translation of core journals, as well as high-quality translations such as foreign propaganda copy and authoritative translations. Ensuring the traceability and content professionalism of the corpus from the source.

3.3 Data Preprocessing

Data preprocessing is mainly text processing. First, the text format should be unified, and data such as special characters and redundant information should be deleted. Next, Chinese needs to complete the conversion of traditional and simplified texts, word segmentation and part-of-speech annotation, and English needs to complete stemming extraction and part-of-speech annotation. At the same time, named entities such as intangible cultural heritage items, inheritors, and technical schools should also be identified and labeled. This ensures the standardization and availability of the text corpus, laying the foundation for subsequent bilingual alignment work.

3.4 Bilingual Parallel Alignment

Alignment work is carried out using the mode of rule alignment and manual calibration. First of all, according to the fixed expressions such as the name of the intangible cultural heritage item and the name of the inheritor, the alignment rules should be formulated. Then, with the help of text features such as chapter titles and paragraph structures, it achieves precise alignment at the text level to achieve phrase-level accuracy. For terms with multiple standard translations, multiple translations should be aligned side by side, and the applicable scenarios of each translation should be marked. All operating results must be manually calibrated by professionals. Deviations in machine alignment are corrected to ensure accuracy and standardization of terminology and translation alignment.

3.5 Storage and Management

Choose the right database as the core database management system. Create a Chinese glossary, an English translation table, a labeling information table, and an alignment relationship table. Through unique identifiers, the data association of each data table is realized, so as to ensure that terms, translations, and annotation information can be accurately corresponded, and to ensure that the text corpus can be accessed efficiently. At the same time, in order to improve the search speed, multiple search indexes are established for terms and keywords and naming core information to meet everyone's needs for quick search terms.

In addition, a sound database optimization and management mechanism is formulated. Regularly back up data, dynamically update corpus, manage user permissions, etc., to ensure stable system operation. Establish a professional expert review process. The new text corpus and translation must be reviewed by intangible cultural heritage experts and translation experts. Through a series of operations, the quality and authority of the corpus are guaranteed from the process.

4. Application of ICH Terminology Corpus

4.1 Applications in the Field of Education

The Chinese-English parallel corpus of intangible cultural heritage terms can provide accurate resource support for translation teaching. Promote students' language learning and ability development; promote teachers' teaching improvement and curriculum construction; Assist in academic research and educational innovation (Liu Zheng, 2025).

Teachers can design translation practice courses, combined with the bilingual content of intangible cultural heritage terms, so that students can intuitively understand the cultural connotation and application scenarios of terms. At the same time, personalized learning paths can also be planned for students at different learning levels. This can not only break the limitations of traditional empiricist teaching, but also meet students' personalized learning needs and solve the problem of "multiple translations of one word". Cultivate scientific and technological translation talents with both language ability and data literacy (Li Rui & Liu Jimeng, 2025).

4.2 Academic Research Application

The application of corpora to English text translation can not only enrich the theoretical system of translation research, but also provide strong support for translation practice (Wu Ye, 2025). The corpus provides rich data support for the translation research of intangible cultural heritage.

Based on bilingual parallel text data, researchers can deeply analyze the linguistic characteristics and translation laws of different intangible cultural heritage terms, and summarize the cross-cultural translation strategies of culturally loaded words. It is also possible to study the international dissemination and acceptance of intangible cultural heritage by comparing Chinese and foreign corpora, and analyze the international dissemination effect of different translation methods. In this way, the research on intangible cultural heritage translation can be promoted from a single case analysis to a systematic, empirical and interdisciplinary direction. Finally, it enriches the theoretical system of cultural translation and terminology translation.

4.3 Cultural Dissemination and Display Application

The corpus provides terminological support for the international dissemination of intangible cultural heritage. We can rely on text resources to build digital platforms such as intangible cultural heritage international communication websites and mobile applications. Publish multilingual terminology comparison, intangible cultural heritage introduction, etc.

Tell the world the story of Chinese intangible cultural heritage through translation. In international exhibitions on the theme of intangible cultural heritage, the text resources of the corpus can be used as the basis for exhibition explanations and graphic introductions, enriching the cultural connotation and educational significance of the exhibition. At the same time, the corpus provides professional terminology reference for the writing of external propaganda copyright and the production of bilingual commentary in the exhibition hall. Avoid the problem of confusion in translation and enhance the international recognition and dissemination of Chinese intangible cultural heritage.

5. Conclusion

Combined with the needs of intangible cultural heritage culture for external dissemination, this study sorts out the core background of the construction of Chinese-English parallel corpus of intangible cultural heritage terms, analyzes the problems including the current lag in cross-cultural communication and foreign translation research of intangible cultural heritage, and the lack of corpus construction related to intangible cultural heritage, and constructs a standardized corpus construction process.

From the three core dimensions of education and teaching, academic research, cultural communication and display, this paper puts forward the application of corpus combined with the needs of external communication of intangible cultural heritage culture, sorts out the core background of the construction of Chinese-English parallel corpus of intangible cultural heritage terms, analyzes the problems including the current lag in cross-cultural communication and foreign translation research of intangible cultural heritage, and the lack of corpus construction related to intangible cultural heritage, and constructs a standardized corpus construction process. Effectively solve the current problems such as irregular English translation of intangible cultural heritage terms and deviations in cultural dissemination. Provide professional support for the translation and introduction of intangible cultural heritage to foreign countries, and promote the standardized development of the translation of intangible cultural heritage terms. At the same time, it provides resources for translation teaching and academic research to help cultivate compound translation talents.

However, the corpus construction plan proposed in this study still has some shortcomings in terms of the coverage of intangible cultural heritage terms with local characteristics. In the future, we need to carry out further research and explore more targeted cross-cultural translation strategies, so that the corpus can truly become a favorable tool for cross-cultural communication of intangible cultural heritage. Help China's intangible cultural heritage to shine better on the international stage, and promote the international dissemination of Chinese culture and cultural exchanges and mutual learning between China and foreign countries.

Fund Project

This paper is supported by the 2025 Annual Open Fund of Hubei Provincial Research Center for Intangible Cultural Heritage; Project Title: Research on the Construction of a Chinese-English Bilingual Corpus of Chinese Intangible Cultural Heritage Terms; Project No.: HGFY202502001.

References

- Li Rui, & Liu Jinneng. (2025). Application of Corpus Tools in University Science and Technology Translation Teaching. *Overseas English*, 2025(23), 98-100.
- Liu Zheng. (2025). Research on Corpus-Based Approaches in Academic English Teaching. *Overseas English*, 2025(22), 61-65.

- Wu Ye. (2025). Corpus-Based Translation of English Texts. *Jiaying Literature*, 2025(15), 92-94.
- Zhang Li. (2025). Research on the Pathways for Overseas Dissemination of National Intangible Cultural Heritage Based on Parallel Corpora. *English Plaza*, 2025(12), 27-30.
- Zhou Ying. (2018). Research and Application of Constructing Small-Scale English-Chinese Parallel Corpora. *SME Management and Technology* (Late Edition), 2018(07), 92-93.