Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230197

A Corpus and Computer-Assisted Translation-Based Study on English Translation of Intangible Cultural Heritage Terms

Menglei Lv [a,1] and Liping Jiang [b]
[a] *Software Engineering Institute of Guangzhou, Guangzhou city, Guangdong province, China*
[b] *Guangdong Mechanical & Electrical Polytechnic, Guangzhou city, Guangdong, China*

Abstract. The translation of intangible cultural heritage (ICH) texts presents significant challenges due to their rich cultural connotations and unique regional characteristics. To address this, the author of this study utilized Corpus and Computer-assisted Translation (CCAT) to assist in the English translation of ICH during the creation of a parallel corpus. Three patterns of CCAT-based English translation of ICH were proposed: machine translation, human-machine interactive translation (combining human and machine translation), and human translation. The implementation of these patterns can improve the efficiency and quality of ICH translation, promote terminology standardization, and facilitate the exportation and transmission of ICH.

Keywords: Corpus and Computer-assisted Translation (CCAT); Intangible cultural heritage; Chinese-English Translation; Intangible cultural heritage terms

1. Introduction

China's rich and extensive culture has a history of thousands of years, and with the implementation of the "the Belt and Road" initiative, promoting cultural "going global" has become an excellent opportunity for cultural development and dissemination [1]. Intangible cultural heritage (ICH) is a crucial component of China's traditional culture and its translation serves as a cultural window for foreign exchange. However, due to the distinct regional and dynamic characteristics of ICH, its English translation presents significant challenges [2].

The author uncovered a wide variety of problems in the existing English translation corpus of ICH texts during the creation of a parallel corpus. For example, in the introduction of local cultural ICH, there are phenomena of excessive transliteration. For instance, "英歌" in Jieyang, Guangdong Province was translated as "Yingge" and "雷剧" in Zhanjiang, Guangdong Province was translated as "Leiju". This simple

[1] Corresponding Author, 1531447813@qq.com, Guangdong Mechanical & Electrical Polytechnic, Guangzhou city, Guangdong, China.

transliteration fails to take into account the acceptability of the target language readers and could not accurately convey the cultural connotation of ICH [3]. Among them, the most severe and influential problem on the international dissemination of ICH is the lack of term standardization, namely, the problem of confusion in the translation of names. The lack of uniformity in translated names in ICH texts indicates the translators' failure in adopting modern translation technology to standardize their translation and also implies low translation efficiency [4].

Machine translation continuously updates the processing paths and means for natural language. The introduction of artificial intelligence has induced machine translation into the era of neural network translation, greatly improving translation speed and efficiency [5]; Applying the computer-assisted translation system to E-C translation and C-E translation has advanced the overall quality to a passing level or above, and enables users to basically achieve the communicative purpose [6]; The construction of a bilingual parallel corpus has provided the best platform for language research and translation research [7]. By constructing and applying a corpus for ICH translation research, researchers can more deeply explore and recognize the laws and characteristics of ICH texts at the linguistic level, and hence summarize objective and universal principles and methods for ICH translation [2]. Using modern translation techniques, including corpus, translators can not only achieve terminology unification, but also greatly improve the accuracy and efficiency of translation [4]. Corpus and Computer-Assisted Translation (CCAT) [1] provides an excellent way for translators to improve the efficiency of intangible cultural heritage translation.

Based on the theoretical concept of CCAT (Corpus and Computer-Assisted Translation), this research takes the English translation of introduction text of the national ICH as an application case to explore the English translation pattern of ICH under the CCAT platform. The corpus used is a national Chinese-English (C-E) bilingual corpus of ICH, which contains three million aligned words of ICH corpus, covering national ICH corpus texts from more than ten cities across the country [2]. Translators can search the corpus for authentic translation examples to assist in translation practice, and can also use translation assistance platforms, such as Snowman CAT software and Trados software, to assist their translation, aiming to achieve the goal of unifying terminology and improving translation efficiency.

2. Literature review

Corpus and Computer-Assisted Translation (CCAT), in which Corpus mainly refers to bilingual corpus, mainly specialized in bilingual corpus and includes online corpus as well, while the computer-assisted part specifically refers to CAT software [1]. The future CCAT system will be built as an online platform and connected to idioms, henceforth every user of the system, such as the translator, the proofreader, the reviewer, and the translation researcher, will belong to the sub platform of the system, sharing the public corpus, terminology, and memory database [4].

Laviosa [8] pointed out that corpus research has been thoroughly integrated into translation research since the early 1990s, and with the development of corpus research, it has also exerted a significant impact on the cultivation of our translation thinking abilities and translation teaching. The symposium on "Corpus Use and Learning to Translate" held in Italy in 1997 opened a new chapter in the application of parallel corpus to interpreter training/translation teaching [9]; Since then, foreign scholars such as Zanettin (1998:616-630) [10], Bowker (1998:631-651, 2003:169-183) [11][12], Barlow

(2000:106-115) [13], Bernardini (2004a: 15-34, 2004b: 97-112) [14][15], and Pearson (2007:15-24) [16] have begun to study the application of corpus in translation teaching and make teaching attempts. Corpus based translation research started relatively late in China, but it has developed rapidly. In 2003, Ye (2003:41-44) [17] explored the issue of C-E translation at the lexical level by building an English Chinese parallel corpus. Wang (2004:27-32) [18] is recognized as a pioneer in exploring the application of corpus in translation teaching, initiating a boom in domestic attempts to use corpus in translation teaching.

The application of corpus and computer-assisted translation systems into translation research has caused heated debate in the translation community. The author conducted a search on CNKI and the result is listed as follows: Based on "Corpus" and "Translation", 9556 results has been shown; Based on "Computer-Assisted" and "Translation", 2282 results has been displayed; Based on "Corpus and Computer-Assisted Translation (CCAT)", 254 results has been exposed. For example, Shen and Zhang (2018:254) [19] took the snowman CAT software as an example to explore the application of computer-assisted translation software based on parallel corpus in translation teaching. They constructed a comparative experimental model and elaborated its application in translation teaching, and then proposed an optimized innovation model based on the results of the experiment; Zhu and Chen (2015:52) [20] developed a computer-assisted translation teaching platform and applied it in translation practice teaching. According to visualized research data, few studies have combined computer-aided translation with the ICH translation pattern for research. However, translation research and teaching applications based on corpus and computer-assisted translation software have become a research trend, so it has certain exploration significance applying corpus and computer-assisted translation in ICH translation research.

3. Three patterns of the CCAT-based English translation of ICH

CAT Translation Memory technology stores and classifies diverse and massive language data in the form of a technical corpus and corpus and then implement the automatic matching during translation. Meanwhile, the main popular CAT tools, such as Trados and DejaVu, also provide the function of convenient text content capture [21]. CAT software mainly has the following three functions: 1) Managing the translation project, facilitating terminology standardization. 2) Avoiding repetitive translation thus to improve the translation efficiency. 3) Establishing a matching mechanism hence to improve the translation efficiency.

Based on the self-built national ICH corpus, utilizing existing original texts and translations, the research team has formed a translation memory and then updated it in real time. When the translator is translating ICH, the computer-assisted translation system will automatically search and match the same or similar translation resources (such as phrases and sentences) in the translation memory, and then provide reference for translators. Therefore, translators only need to focus on the translation of new content, which can not only avoid unnecessary repetitive work, but also can ensure the unity of the translation of the same content.

Owing to the fact that national ICH translation materials contain amounts of expressions with rich cultural connotations, dealing with culture-loaded words simply with CAT model has revealed numerous drawbacks. Therefore, this study will propose three patterns of the CCAT-based English translation of ICH: machine translation, human-machine interactive translation (combining human and machine translation:

human-assisted translation and machine-assisted translation), and human translation, based on the terms and special expressions with rich cultural elements in ICH texts.

3.1. Machine translation

Machine translation has a relatively high translation speed and efficiency, and has been widely used in translation fields with high degree of informative texts. Translation software can be adopted directly to translate simple and objective declarative sentences with little cultural connotation and plain language in some ICH texts.

Example 1: 五羊传说(from national level ICH of Guangdong)

CAT: The legend of the five rams

The English translation shown in Example 1 is simple and clear, and can be easily understood by target language readers, so it can be adopted directly. Then, by searching for the English translation of the term "五羊传说" in the network corpus, the author has further verified that the term is professional.

3.2. Human-machine interactive translation

Machine translation and human translation have their own advantages and disadvantages respectively. The advantage of machine translation is its fast speed and high efficiency, but some machine translations may be rigid, obscure, and fail to accurately convey the cultural connotations of ICH. Although human translation is more accurate than machine translation to some extent, it cannot efficiently handle a large amount of translation tasks or those emergent tasks. The human-machine interactive translation pattern combines both translation methods, complementing each other's advantages. By making full use of parallel corpora of ICH, network corpora, and computer-assisted translation systems, translators can conduct human translation to maximize translation effectiveness.

The human-machine interactive translation includes two main operation modes: human-assisted translation and machine-assisted translation.

3.2.1 Human-assisted translation

With today's technological advancements, machine translation still cannot completely replace human translation. To achieve a balance between translation quality and efficiency, human-assisted translation is necessary. Human-assisted translation should be adopted to translate ICH texts with complex sentence structures and little cultural connotation but simple and objective declarative sentences. Human-assisted translation mainly involves text processing before translation and post-editing after translation. Text processing before translation includes adding a subject to sentences without one based on context, removing redundant vocabulary, and replacing special vocabulary. Post-editing includes selecting and replacing words in the translated text and arranging sentence structure, etc.

Example 2: 历史上由唐代宫廷狮子舞里脱胎而来,五代十国之后,随着中原移民的南迁,舞狮文化传入岭南地区.(National level ICH)

Text processing before translation: A subject has to be added to the first sentence according to the context for there is no subject in it; The chunk "历史上" has been removed for the similar connotations of it with the other chunk "脱胎而来", which was

then replaced by another chunk "演变而来" for the later one is more easily been read by the machine. Thus the original text has been edited and translated as:

CAT: It evolved from the palace lion dance in the Tang Dynasty. After five dynasties and ten countries, with the migration of immigrants from the Central Plains to the south, the lion dance culture was introduced into Lingnan.

Post-editing: As is recommended in the network corpora, it is more acceptable to translate "宫廷狮子舞" as "court lion dance". According to Chinese history, it is more appropriate to translate "五代十国" as "the Five Dynasties and Ten States (907 B. C.~960 B. C.)". Firstly, "五代十国" is a proper noun, hence it should be capitalized. Secondly, according to Chinese history, it is not agreeable to translate "十国" as "ten countries". Furthermore, by searching into the network corpora, the author has found that in 80% corpus, "五代十国" is translated as "Five Dynasties and Ten States". Finally, adding the time annotation after the expression "五代十国" can help the target language readers to understand the history and culture during such a period. Moreover, the sentence structure has been adjusted due to the different sentence structures of Chinese and English.

Post-edited version: It evolved from the court lion dance of the Tang Dynasty. After Five Dynasties and Ten States (907 B. C.~960 B. C.), the culture of lion dance was introduced into Lingnan as the people in the Central Plain immigrated to the south.

3.2.2 Machine-assisted translation

ICH texts contain massive terms with cultural connotations, thus, as previously mentioned, English translation of ICH terms is deemed as a difficulty. For example, there are phenomena of excessive transliteration. For instance, "佛山十番" was translated as "Foshan Ten Scenes" and "英歌" of Jieyang was translated as "Jieyang Yingge". Such simple transliteration fails to consider the acceptance of target language readers, and cannot accurately convey the cultural connotations of ICH. While machine translation of these terms is also not ideal. According to the literature, "佛山十番" is a type of folk instrumental music, which belongs to gong and drum music. Therefore, "Foshan Percussion Music" is a more appropriate translation. As is recorded in the literature, there are several types of "英歌", including "英歌" (Puning, Jieyang), "潮阳英歌" (Shantou), and "甲子英歌" (Shanwei), all of which have different characteristics for they originated from different regions. Therefore, the adjusted translations of "Puning Ying-Ge Dance", "Chaoyang Ying-Ge Dance", and "Jiazi Ying-Ge Dance" are more appropriate and easier for target language readers to accept.

To ensure the terminology standardization, avoid repetitive translation of the same content, save resources, and ensure the accuracy of translations, translators can adopt machine-assisted translation with the help of computer-assisted translation (CAT) tools, ICH parallel corpora, online translation, and translation dictionaries. Translators can verify their translations of terms via the network materials and Chinese-English parallel corpora, and then added the verified terms to the ICH Chinese-English corpus for real-time updating, hence to standardize the translation of ICH terms through the translation memory retrieval and matching.

3.3 Human translation

Human translation can be flexibly adapted to specific contexts and grammatical requirements. For special expressions with distinct cultural connotations, human

creativity in translation can make it easier for the target readers to understand. Based on the source content, human translators can create new expressions in the target language by combining transliteration and free translation, which can not only help the readers understand the source content better but also have a deeper impact on the readers. In fact, it is more conducive to the dissemination of Chinese culture.

Example 4: 主要套路有“采青”, “高台饮水”, “狮子吐球”, “踩梅花桩”等. (National level ICH)

CAT: The main routines include “picking green”, “drinking water from high platform”, “lion spits ball”, “stepping on plum blossom pile”, etc.

The original text contains numerous expressions with rich cultural connotations. Machine translation mostly adopts the method of free translation to deal with these expressions, and the part of speech of theses expressions in the machine translation are not uniform. By searching the relevant information on the internet and then studying the translations in the ICH corpus, the author has confirmed the translations of the terms and then stored them into the translation memory to update the ICH translation corpus. Additionally, the CAT translation reads still too rigid, thus it is adjusted as follows:

Human translation: They will also represent the set patterns of “Caiqing” (picking the green), “Gaotaiyinshui” (drinking on a hathpace), “Shizituqiu” (spitting small balls) and “Caimeihuazhuang” (stepping on the staggered pilings).

In terms of the obscure Classical Chinese rich in the complex traditional Chinese cultural elements, human translation is an effective translation strategy.

4. Conclusion

In conclusion, this paper highlights the benefits of using modern translation technologies, such as corpora and computer-assisted translation systems, on the CCAT platform to integrate human and machine translation for the English translation of ICH. The proposed translation patterns, including machine translation, human-machine interactive translation (combining human and machine translation), and human translation, provide effective solutions for different types of ICH texts. The study demonstrates that the CCAT-based approach can significantly improve translation efficiency and accuracy, ensuring the unification of terminology in translations.

Moving forward, this research can be applied to other translation fields that require creative translation of culture-loaded words, ambiguous language, and professional terms. The next step of the research is to apply the ICH corpus in translation teaching and construct an ICH English translation teaching platform system under the CCAT platform. The proposed new translation teaching modes will be tested through empirical research on translation major students, offering valuable insights and guidance for research on translation teaching in the new era.

Acknowledgement

This research was supported by the Humanities and Social Sciences Research and Planning Fund Project of the Ministry of Education of China (2021) (Grant No. 21YJA740016); the Key Scientific Research Project (Special Scientific Research Project) from Guangdong Mechanical and Electrical Polytechnic (Grant No. YJZD2021-47).

References

[1] Yu, J. Translation research on world heritage under the CCAT platform. Journal of Qiqihar University (Philosophy, Social Science Edition), 2018,(02), 135-137.

[2] Jiang, L. P. Creation and application of the Chinese-English parallel corpus of intangible cultural heritage. China's Intangible Cultural Heritage, 2022,(02), 118-124.

[3] Jiang L, Al-Shaibani GKS, Yang F, Cheng M and Huang M (2022) The metonymic mechanism of English translation of Chinese intangible cultural heritage terms from the perspective of cognitive psychology. Front. Psychol. 13:957485. doi: 10.3389/fpsyg.2022.957485

[4] Jiang, L. P. Exploring translation strategies of intangible cultural heritage from the perspective of cross-cultural communication: Taking the English translation of Guangdong intangible cultural heritage as an example. Journal of Qingyuan Vocational and Technical College, 2022,15(01), 69-74.

[5] Jiang, L. (2022). Factors influencing EFL teachers' implementation of SPOC-based blended learning in higher vocational colleges in China: A study based on grounded theory, Interactive Learning Environments, DOI: 10.1080/10494820.2022.2100428

[6] Yang, W. D., & Fan, Z. R. Analysis of post-editing in machine translation of scientific and technological texts. Shanghai Translation, 2021,(06), 54-59.

[7] Li, F. Q. A comparative study on the quality of English-Chinese machine translation based on neural networks. Shanghai Translation, 2021,(04), 46-52.

[8] Wang, K. F. . The design and development of Chinese-English parallel corpus. China Foreign Language, 2012,9(06), 23-27.

[9] Laviosa, S. Corpus-based translation studies: theory, findings, applications. Amsterdam: Editions Rodopi B.V.,2002.

[10] Xiong, B. Research on translation teaching mode based on Chinese-English parallel corpus. Foreign Language World, 2015,(04), 2-10.

[11] Zanettin, F. Bilingual comparable corpora and the training of translators. Meta, 1998, 43(4), 616-630.

[12] Bowker, L. Using specialized monolingual native-language corpora as a translation resource: A pilot study. Meta, 1998, 43(4), 631-651.

[13] Jiang, L., Liang, X. (2023). Influencing factors of Chinese EFL students' continuance learning intention in SPOC-based blended learning environment. Educ Inf Technol. https://doi.org/10.1007/s10639-023-11734-4.

[14] Barlow, M. Parallel texts in language teaching. In Botley, S.P., McEnery, A.M., & Wilson, A. (Eds.), Multi-lingual Corpora in Teaching and Research (pp. 106-115). Amsterdam: Rodopi Bv Editions,2000.

[15] Bernardini, S. Corpora in the classroom: An overview and some reflections on future developments. In Sinclair, J.M. (Ed.), How to Use Corpora in Language Teaching (pp. 15-34). Philadelphia: John Benjamins,2004.

[16] Bernardini, S. Corpus-aided language pedagogy for translator education. In Malmjær, K. (Ed.), Translation in Undergraduate Degree Programmes (pp. 97-112). Amsterdam: John Benjamins,2004b.

[17] Pearson, J. Using parallel texts in the translator training environment. In Zanettin, F., Bernardini, S., & Stewart, D. (Eds.), Corpora in Translator Education (pp. 15-24). Beijing: Foreign Language Teaching and Research Press,2007.

[18] Ye, C. Q. The application of self-built corpus in translation teaching: Classroom design for the translation of the Chinese and English versions of "A Dream of Red Mansions". Foreign Languages and Their Teaching,2003, (03), 41-44. DOI: 10.19716/j.1672-4720.2003.03.009.

[19] Wang, K. F. The use of bilingual parallel corpus in translation teaching. Foreign Language Electronic Teaching, 2004,(06), 27-32.

[20] Shen, Y., & Zhang, L. The application of computer-aided translation software based on parallel corpus in translation teaching: Taking the XueRen CAT software as an example. In Research on the Development and Innovation of Foreign Language Education and Translation 2018,(Vol. 7, pp. 254-257).

[21] Zhu, Y. B., & Chen, J. L. (2015). Theoretical conception and engineering practice of computer-aided translation teaching platform based on corpus. Foreign Language Electronic Teaching, (04), 52-57.

[22] Zhao, H. (2019). English translation of cross-border e-commerce product copywriting with human-machine interaction. Chinese Science & Technology Translators Journal, 32(01), 26-28+8.