A Corpus-Based Study on Nanchang Red Tourism Translation under the Background of the "100th Anniversary of the Founding of the Communist Party of China"

Yuan Liu

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

Abstract: Under the background of the 100th anniversary of the founding of the Communist Party of China, this study adopts the English translation of Nanchang red tourism and the English text of foreign red tourism as two sub-corpuses of a self-built comparable corpus based on the corpus and linguistic methods. The results of the comparative study are used to propose better strategies for the translation of Nanchang red tourism texts, to come up with the most suitable method for the translation of Nanchang red tourism, and to provide reference for its subsequent translation and the English translation of red tourism in China.

1. Introduction

In 1993, Mona Baker, a British scholar, published the article "Corpus Linguistics and Translation Studies: Implications and Applications", in which Baker pointed out that the turning point in the direct consequence of access to large corpora of both original and translated texts, and the development of specific methods and tools for interrogating such corpora in ways which are appropriate to the needs of translation scholars [1]. Subsequently, Corpus Linguistics began to attract the attention of more translation researchers, and the corpus-based red tourism translation research is an important application of it. This year is the 100th anniversary of the founding of the Communist Party of China, which is of great significance for the development of red tourism. While Nanchang, as the cradle of China's red revolution, enjoys rich red tourism resources and attracts a large number of domestic and foreign tourists every year. Under the background of the 100th anniversary of the founding of the Communist Party of China, improving the translation of Nanchang red tourism not only produces important significance in terms of foreign publicity and investment attraction, but also a powerful way to enhance the international image of Nanchang. Due to various factors, there are still many shortcomings in the English translation of Nanchang red tourism texts, and many large tourism text corpuses are not open to the public, which makes it difficult to conduct corpus-based red tourism translation study.

2. Study Methods

There are three main ways to study the English-Chinese tourism text corpus: The first is Comparable Corpus of Original English and Chinese Tourism Texts, which includes two sub-corpuses: Comparable Corpus of Original English Tourism Texts and Comparable Corpus of Original Chinese Tourism Texts, and there is no translation relationship between them; the second is Comparable Corpus of English Tourism Texts, which includes English Sub-corpus and Translated Sub-corpus; the third is Parallel Corpus of Chinese-English Tourism Texts Parallel Corpus of Chinese-English Tourism, which includes two sub-corpuses: Sub-corpus of Source Language and Sub-corpus of Target Language, and the two sub-corpus are in translation relationship [2]. Based on the current collection of Nanchang red tourism texts and foreign red tourism texts, the results of the experimental data are compared with the comparable English corpus, and the differences between the

English translation of Nanchang red tourism texts and foreign red tourism texts are further understood.

2.1 Sources and Processing of the Corpus

In order to explore and improve the translation of Nanchang red tourism texts, I have built a comparable corpus, which contains two sub-corpuses: the Translated English Red Tourism Text Corpus (TERTT) and the Original English Red Tourism Text Corpus (OERTT). English Red Tourism Text Corpus (OERTT). The text collection of the corpus is mainly from the following aspects: Nanchang red tourism books, paper versions of publicity pages, field photography of the introduction of Nanchang red tourist attractions, and publicity texts and explanatory words on Nanchang tourism websites, while the English source language corpus is mainly from books related to the theme of red revolution in foreign countries, as well as English electronic texts and documents of red revolution on foreign websites.

2.2 Corpus Analysis Tools

There are various standard problems in the texts obtained through the Internet, manual input and scanning, such as symbols and formatting, so the author uses CorpusWordParser, a corpus software developed by the corpus team of Beijing Foreign Studies University, to process the collected corpus to obtain a well-formatted corpus. After the corpus was collected, organized and processed, the English corpus of both corpora was assigned with the TreeTagger_Lite_English developed by the corpus research team of Beijing Foreign Studies University, and after the assignment was completed, the corpus search tool was used for the next step of the study. In this study, I used AntConc, a corpus retrieval tool, according to the suitability of the research content and direction.

3. Study Process

The study process of corpus generally includes data capture, tokenization, and POS tagging.

3.1 Data Capture

In order to ensure the accuracy of the research results, the collected corpus needs to be relevant. At the same time, according to the requirements of the corpus search tool AntConc, all corpus texts should be stored in a uniform format of .txt.

3.2 Tokenization

After the collection of Chinese and English texts, word separation is needed, because the accuracy of the results of this study will be greatly affected by the lack of word separation or the simple coding conversion, such as garbled codes and long words and phrases. Therefore, I used CorpusWordParser, a professional word splitting software developed by the corpus team of Beijing Foreign Studies University, to split the text of the corpus, and then imported the split Chinese and English texts into AntConc3.5.8 to build a Chinese and English corpus for the next step of the study.

3.3 Pos Tagging

Since AntConc does not directly count real words, other tools are needed to assign codes to word categories. The two software programs were able to recognize each word in the Chinese and English texts respectively, and automatically mark the lexicality after the word. The assigned text was then imported into AntConc 3.5.8, and statistics were found according to the assigned values of nouns, verbs, adjectives and adverbs.

4. Vocabulary Feature Data Collection and Analysis

The overall statistical features of corpus texts usually include: bytes, tokens, types, Type-token ratio (TTR), and lexical density. The value of class/form ratio is proportional to the lexical richness of the text [3]. By importing the texts into the corpus analysis software AntConc 3.5.8, the specific data results are shown in Table 1 below.

Table 1 Basic Data

	OERTT	TERTT
Tokens	30345	32486
Types	3945	4044
TTR	0.13	0.124

The statistical results in Table 1 show that OERTT has a higher TTR than TERTT, so the degree of lexical variation and richness in OERTT is greater than TERTT, indicating that the foreign red tourism texts are richer than the English translation of Nanchang tourism in vocabulary.

4.1 Lexical Density

Lexical density: the average value of content words (nouns, verbs, adjectives, adverbs) in a clause. Its main function is to illustrate the content of the discourse by analyzing the ratio of the number of content words to the total number of words.[4]. Jcan ure proposed the formula for calculating lexical density: lexical density equals to the number of content words divided by the total number of words multiplied by 100%. It is also pointed out that lexical density is a criterion to distinguish the degree of formality of a text. The more real words in the text, the higher the lexical density, the more formal the text is and the richer the information conveyed; on the contrary, the text is relatively less formal and conveys less information [5]. The specific data results are shown in Table 2 below.

Table 2 Lexical Density Statistics

Corpus	Nouns, Verbs, Adjectives, Adverbs	Total content words	Size	Lexcial density
OERTT	10455,2950,2496,950	16851	30345	55.5%
TERTT	12449,3672,3416,996	20553	32486	63.3%

By observing Table 2, we can find that the lexical densities of the OERTT and TERTT are 41.8% and 52.5% respectively, and it is obvious that the lexical density of the former is less than that of the latter, which also indicates that TERTT has a higher proportion of content words and conveys a richer amount of information, and its context is also more difficult. The number of nouns, adjectives and adverbs in TERTT is 12,449, 3,416 and 996 respectively, which is much larger than that of OERTT (10,455, 2,496 and 950), and the number of verbs used in TERTT is slightly larger than that of OERTT, because in Chinese, it is common to use a combination of multiple verbs in a chronological and factual reasoning way [6]. It can be seen that the translators who translated this Nanchang red tourism text were influenced by some negative transfer from their native language (Chinese) (due to the cultural differences, non-native speakers are influenced by some language rules and expression habits from their native language when they learn and use a second foreign language). At the same time, the comparison of the amount of adjectives and adverbs used shows that the English translation of Nanchang red tourism text is more vivid and beautiful, and the style is more written than the foreign red tourism texts, which is inseparable from the rich red cultural and historical heritage of Nanchang.

4.2 Personal Pronouns

Personal pronouns in English include the following three kinds: first person (I/we/my/our), second person (you/your excluding the object you) and third person (he/she/it/they/his/her/its/their excluding the object it). By looking at the search results of AntConc 3.5.8, the statistics of personal pronouns in OERTT and TERTT can be derived, as shown in Table 3.

Table 3 Personal Pronouns Statistics

Corpus	First person	Second person	Third person
OERTT	97	23	187
TERTT	71	14	257

From Table 3, it can be seen that the number of first and second persons used in OERTT is higher than that of TERTT, 1.3 times and 1.6 times respectively, while the number of third person pronouns is lower than that of TERTT. The more frequent occurrence of first and second persons in OERTT shows that it pays more attention to the authenticity of describing things, and increase the sense of intimacy and inadvertently draw closer to the readers. In contrast, TERTT describes things in an

objective and plain writing style. At the same time, the third person pronouns in two corpuses are even much larger than the sum of the first and second persons, and the third person in OERTT is more than in TERTT, which is analyzed mainly because when referring to information such as places of interest or people, pronouns are usually repeatedly used in TERTT, while OERTT prefers to use original words (non-pronouns).

5. Study Findings

With the help of a series of corpus analysis tools, I conducted a statistical analysis of the data results of a self-built comparable corpus of red tourism translations and summarized the following conclusions: firstly, by comparing the data of the TTR of the two corpora, the degree of lexical variation and richness in OERTT is greater than that in TERTT, indicating that foreign red tourism texts are richer in terms of words and more difficult to read than the English translation of Nanchang red tourism texts. Secondly, the statistical analysis of word distribution shows that there are more content words in TERTT than OERTT, which shows that the English translation of Nanchang red tourism is richer in information and more formal in style, which is inseparable from the strong red cultural heritage of Nanchang. In terms of lexical density, TERTT holds a higher indicator, which shows that native Chinese speakers are more influenced by the negative migration of their mother tongue when translating red tourism texts, and tend to use the syntactic structure of Chinese to express themselves, which is not in line with the expression habits of English. At the same time, the number of first and second persons in OERTT is more frequent, which shows that it focuses more on the authenticity of describing things, increase the sense of intimacy and inadvertently draw closer to the readers, while TERTT describes things in an objective and plain writing style.

6. Conclusion

Based on the self-built comparable corpus of red tourism translation, this study compares and analyzes the linguistic features of foreign red tourism texts and Nanchang red tourism texts, with the aim of identifying the shortcomings in the English translation of Nanchang red tourism with reference to foreign red tourism English texts, so as to improve the translation quality of English translation texts and make the language expression conform to the English text specification. Meanwhile, under the background of the "100th anniversary of the founding of the Communist Party of China", we will vigorously promote the international publicity of red tourism in Nanchang. The corpus collected by the author is not large enough, so I expect to conduct a more in-depth and comprehensive study on the original basis in the future, in order to provide references for the relevant departments and units of Jiangxi Province in decision-making and publicity, better spread the red tourism culture and history of Nanchang to foreign countries, attract more foreign tourists to Nanchang and enjoy the charm of local red tourism culture, and thus accelerating the promotion of Nanchang red tourism to the international level.

7. Acknowledgment

References

[1] Baker M. Corpus Linguistics and Translation Studies: Implications and Applications[M]. M Baker. G Francis, E Tognini-Bonelli(eds). Text and Technology: In Honour of John Sinclair. Amsterdam: John Benjamins, 1993: 235

[2] Xiaopeng Liang, Ning Kang. On corpus-based approach to the translation of Chinese tourism texts [J].Journal of Qingdao University of Science and Technology,2010,26(04):115-117+120.

[3] Baker M. Towards a methodology for investigating the style of a literary translator [J]. Target, 2001, 12(2):241-266.

[4] Xinzhang Yang. Features of lexical density in written English corpus[J].Journal of PLA University of Foreign Languages,1995(03):14-18.

[5] Halliday, M.A.K. Spoken and Written Language[M]. New York: Oxford University Press. 1985: 88.

[6] Xiaoyu Deng, Zhenping Wang. A Cultural Perspective on Translation Theory [M].Changchun:Jilin University Press,2015:110.