Retraction

Retraction: A Corpus-based Linguistic Analysis of Translated Tourism Interpretive Panels under the Background of “Big Data” (IOP Conf. Ser.: Earth Environ. Sci. 693 012061)

Published 23 September 2022

This article has been retracted by IOP Publishing following an allegation that raises concerns this article may have been created, manipulated, and/or sold by a commercial entity. In addition, IOP Publishing has seen no evidence that reliable peer review was conducted on this article, despite the clear standards expected of and communicated to conference organisers.

The authors of the article have been given opportunity to present evidence that they were the original and genuine creators of the work, however at the time of publication of this notice, IOP Publishing has not received any response. IOP Publishing has analysed the article and agrees there are enough indicators to cause serious doubts over the legitimacy of the work and agree this article should be retracted. The authors are encouraged to contact IOP Publishing Limited if they have any comments on this retraction.

Retraction published: 23 September 2022
A Corpus-based Linguistic Analysis of Translated Tourism Interpretive Panels under the Background of “Big Data”

Jing Xiong*

Department of Foreign Languages, Nanchang Institute of Technology, TianXiang Road No.289, Hi-tech District, Nanchang, China.
Email: xiongjing0792@163.com

Abstract. This paper collects the interpretive panels from scenic spots in China and abroad and then establish two corpora (“CTETIPC” and “CETIPA”). It analyzes and compares the linguistic features of translated and original English tourism interpretive panels from the aspects of STTR, MWL/MSL, nominalization, tense and voice. It is concluded that the translated English tourism interpretive panels are in some aspects non-standard. It is advised to take into consideration the different way of English expressions so as to better achieve the communicative effects.

1. Introduction
China has become the fourth largest inbound travel reception country in the world and tourism industry has great significance to the development of national economy. Therefore, it is important to optimize the tourism development environment and improve the tourism translation quality. Tourism interpretive panels can offer information of the scenic spot in the form of pictures and texts. Corpus is an electronic text database based on massive sampling and processing. Aided by computer tools, corpus can be used to conduct theoretical and applied language research.

2. Corpus and Translation
Corpus linguistics is to analyze a large number of language facts from a macroscopic perspective and summarize the rules of language use based on the true language data.[1] The purpose of corpus is to tell us what kind of language type is. Hunston[2] thinks that it is more reliable to rely on the corpus than on the tuition of native speakers in that language in the corpus covers many forms and large varieties. Therefore, it is significant to summarize the language characteristics through the corpus.

In the 1990s, the Translation and Cross-cultural Research Center in Manchester University began to apply the corpus linguistics methodology in translation research and aimed to establish an English corpus. There are following two purposes of adopting the corpus-based method to analyze the translated text. One is to find out whether the characteristics of translated texts including language pattern, language preference, communicative purpose are different from the characteristics of English text written by native speakers. The other purpose is to study the the “translation universals”which is the focus of empiricism.

3. Corpus Description
All of the texts in the corpus are interpretive panels collected from the major scenic spots in China which is written in Chinese and English. After the correction and denoising, clean texts can be obtained in the form of TXT(UTF-8). On this basis, a corpus is founded named “Corpus of Translated English Tourism Interpretive Panels in China” (CTETIPC) with the capacity of 25541 words. In order
to investigate the differences between the translated interpretive panels in China and the original interpretive panels abroad, the writer collected various original interpretive panels from English-speaking countries and created a corpus named “Corpus of English Tourism Interpretive Panels Abroad” (CETIPA) with the capacity of 23752 words. In order to achieve a more scientific comparison, the two corpora should be approximately the same in size and capacity and cover both of natural attractions and cultural attractions.

4. Research Tools

Text processing is needed before database creation. First, Editplus is adopted to delete unnecessary spaces, carriage-returns. After creating the corpus, it is needed to extract the information from the corpus, which is called retrieval. Retrieval tools adopted in this research are wordsmith5.0 and Antconc which can not only search for the specific node words but also those co-occurring words and make a list of them. Regular expression makes it possible to conduct a precise and strict retrieval mission in the aspects of segment length and structure, which is helpful to improve the accuracy and efficiency of retrieval.[3]

5. Contrastive Analysis of Corpora

5.1. TTR and MWL/MSL

The term Token means all of the words in a text, while the term Type refers to the number of unduplicated words in the text. So TTR (type/token ratio) is generally used to analyze the lexical richness. The larger the ratio is, the richer lexical resource a text contains. But different corpus has different capacities, so only when we standardize the ratio, which means we calculate the ratio on the basis of every a thousand words and draw an average value, which is called the standard type/token ratio (STTR), can the ratio bear a reference value.[4]

Mean Word Length means the average number of syllables or letters a word in a corpus has. Mean Sentence Length refers to the average number of words a sentence in a corpus has. In general, the shorter the Mean World Length and the Mean Sentence Length is, the more readable a text will be, and the longer the Mean World Length and the Mean Sentence Length is, the more difficult a text will be.

Table 1. TTR of CTETIPC & CETIPA

|        | CTETIPC | CETIPA |
|--------|---------|--------|
| Token  | 25541   | 23752  |
| Type   | 2243    | 2389   |
| TTR    | 8.78    | 10.1   |
| STTR   | 39.76   | 41.45  |

Table 2. MWL and MSL of CTETIPC & CETIPA

|        | CTETIPC | CETIPA |
|--------|---------|--------|
| MWL    | 4.23    | 4.62   |
| MSL    | 19.3    | 16.1   |

As is shown in table one, the TTR of CTETIPC is smaller than that of CETIPA, with the former being 8.78 and the latter being 10.1. Since the two corpora differ in capacity, we need to calculate the STTR respectively in order to get a more scientific comparison. After this step, we find that the STTR of CTETIPC (39.76) is still smaller than that of CETIPA (41.45), which illustrates that there is relatively less variation of words in the English translated panels in China’s scenic spots compared to that in foreign scenic spots. And as is shown in chart two, the mean word length in CTETIPC is 4.23, a bit smaller than that in CETIPA, which is 4.62. This result illustrates that the translated English panels in China’s scenic spots have less lexical difficulty and is less formal than original English panels in foreign scenic spots.

Besides, chart two also counts the mean sentence length in CTETIPC and CETIPA, and it shows that the former (19.3) is larger than the latter (16.1), whereas the TTR and the mean word length of
CTETIPC are smaller than that of CETIPA, as mentioned above. Taking into account the gap in the language ability between the translator and the native English speaker, the sentences in the translated English panels in scenic spots are supposed to be simple and have a shorter length and less complexity than those in foreign scenic spots. However, the truth is on the contrary. Influenced by the negative transfer of Chinese grammar, the translated panels in China’s scenic spots adopt a good many long run-on sentences, which may result in a lot of grammatical errors and end up failure in effective communication.

5.2. Nominalization

Nominalization is defined as “the process or result of forming a noun or noun phrase from a clause or a verb”[5]. This paper mainly discuss the application of the nominalization in CTETIPC and CETIPA. Antconc3.5.8 tool is adopted to retrieve the words ended with -ment/-ion/-ness/ which represent “process” and “feature”. Furthermore, it is needed to exclude the disturbance terms such as “moment”, “billion”, “business” and so on, because these words are not transformed from verbs or adjectives. The retrieval is as follows:

|                | CTETIPC | CETIPA |
|----------------|---------|--------|
| -ment          | 412 (1.61%) | 343 (1.44%) |
| -ion           | 97 (0.37%)  | 42 (0.17%)  |
| -ness          | 365 (1.43%) | 281 (1.18%) |
| -ity           | 182 (0.71%) | 132 (0.56%) |

As shown by the table above, the occurrence frequency of nominalization in translated English tourism panels in China’s scenic spots is higher than the original panels in foreign scenic spots. Nominalization can condense the information, simplify the language structure as well as improve the objectivity, because it is unnecessary to present the action performer[6]. On the other hand, it makes the narration too objective and rational, as if it has nothing to do with the reader, which reduces the appeal of the text and enthusiasm of the reader.

5.3. Voice Feature

This study analyzes the voices of translated tourism panels in China and the original panels in foreign countries from the aspect of occurrence frequency. The regular expression is adopted in Antconc3.5.8 to search all the sentence patterns containing “be/get”.

|                | CTETIPC | CETIPA |
|----------------|---------|--------|
| Token          | 25541   | 23752  |
| Number of Passive Voice | 1754    | 1175   |
| Frequency of Passive Voice (Every a thousand words) | 68.7    | 49.5   |

As is shown by the table, the frequency of the passive voice in the translated tourism panels (per a thousand words) in China (68.7) is higher than that in the original English tourism panels(49.5). A lot of active sentences are translated into passive sentences. By digging deep the reason, we can find that it is because that the Chinese grammar is covert, so the subjects in Chinese sentences are always omitted or implied in sentences.[7] When the Chinese sentences are translated into English, it is always necessary to supplement the subject or transform the object in Chinese to the subject in English, thus resulting in the passive voice. Furthermore, Chinese is a “topic-prominent” language while English is a “subject-predicate” language. In English, only the noun and nominal structure can serve as the subjects, and all the subjects can dominate the whole sentences. However, the subject in Chinese is more general, most of which are actually topic. Therefore, in Chinese-English translation, it is unavoidable to adopt the subject-predicate structure to transform the sentence structure, which results that many active voices are translated into passive voice.
5.4. Tense Feature
As we all know, there is variety of tenses in English. To demonstrate the differences between the translated tourism panels in China and the original panels abroad in terms of tense, this study adopts the Regular Expression to retrieve the frequency of each kind of tense, including simple present tense, simple past tense, simple future tense, perfect tense and progressive tense. The results are as follows:

| Tense Feature          | CTETIPC     | CETIPA      |
|------------------------|-------------|-------------|
| simple present tense    | 403 (59.3%) | 562 (71%)   |
| simple past tense       | 145 (21.3%) | 135 (17%)   |
| simple future tense     | 26 (3.8%)   | 32 (4%)     |
| perfect tense           | 74 (10.9%)  | 40 (5.1%)   |
| progressive tense       | 32 (4.7%)   | 23 (2.9%)   |

In general, the simple present tense is the most common tense in English tourism texts, most of which are used to introduce the current state of scenic spots. Simple past tense is the second common tense which is usually used to introduce the history and origin of the scenic spots. The perfect tense is used much less in tourism panels because the panels always describe the scenery and history of the scenic spots rather than their future planning. It is noted that the frequency of perfect tense in CTETIPC(10.9%) is nearly twice as much as that in CETIPA(5.1%) from which we can see the translated tourism panels pay more attention to the effect of past events on the present situation.

6. Conclusion
This paper takes corpus of translated English tourism interpretive panels as the research object and discuss the differences between it and the foreign original tourism panels from the following CTETIPC the translated English tourism interpretive panels are in some aspects non-standard. It is advised to take into consideration the different way of English expressions so as to better achieve the communicative effects.

7. Acknowledgments
2018 Humanities and Social Sciences Research Program of Jiangxi Province. The Construction and Application of Bilingual Corpus of Jiangxi Tourism Resource Under the Background of “Belt and Road Initiative” (YY18220); 2019 Social Science Planning Program of Jiangxi Province: Study on Chinese-English Translation of Chinese Water Culture from the Perspective of Communication (19YY19); 2019 Humanities and Social Sciences Research Program of Jiangxi Province: Establishing Parallel Corpus on the Folk Customs in Jiangxi Intangible Cultural Heritage List.(YY19219).

8. Reference
[1] Huizhong Yang. Introduction to corpus linguistics[M]. Shanghai: Shanghai Foreign Language Education Press,2015.
[2] Hunston, S. Corpora in Applied Linguistics[M]. Cambridge: Cambridge University Press, 2002.
[3] Ning Kang. Corpus-based Stylistic Analysis of Tourism English[J]. Language Teaching and Research, 2017(1):129-136.
[4] Scott, M, Tribble, C. Textual Patterns: Keyword and Corpus Analysis in Language Education[M]. Amsterdam & Philadelphia: John Benjamins Publishing Co., 2006,89-100.
[5] Stubbs, M. Text and Corpus Analysis[M]. Oxford: Blackwell Publishers, 2006.
[6] Laviosa, S. Corpus-based Translation Studies: Theory, Findings and Application[M].Amsterdam: Rodopi,2016: 56-71.
[7] Miqing Liu. English and Chinese Language Contrast and Translation[M].Beijing: China Translation &Publishing Corporation, 2006: 67-85