A Corpus-Based Contrastive Study of Film and TV News: China Daily and Huffington Post as an Example

Yang Pan

School of Foreign Languages, Jiangxi University of Technology, Ganzhou 341000, China.

Abstract: Translating the news about Chinese film and television well can promote the dissemination of Chinese culture. Based on a comparable corpus, this paper explores the differences between the English translation of Chinese film and TV news from China Daily and the English news from Huffington Post at the lexical and syntactic levels. It empirically describes the linguistic features of the news from China Daily. The findings show that, China Daily's film and television news uses more difficult vocabulary and syntax. As a result, it is more difficult to read. In order to introduce China's excellent film and television works to foreigners more effectively, the author gives some suggestions based on the above findings.

Keywords: Comparable Corpus; Film and Television News; Translation; Linguistic Features; Cultural Communication

1. Introduction

News is an important means of transmitting information. Through news, foreigners can understand a country and form a certain impression on it. The quality of news translation will directly affect the spread of national culture, which should be paid attention to. Most film and television news will introduce some interesting and excellent film and television works, which can attract foreign ordinary people to read. These film and television works often contain many elements of Chinese culture, which can enable foreigners to better understand China and change their views on China. Therefore, translating such film and television news is of great significance for the output of Chinese culture. In fact, the English translation of Chinese film and television news is mostly done by Chinese translators. Is there any significant difference between these news and that written by American journalists? What are the differences? Both are worth exploring.

In recent years, there have been many studies on English news translation in China, but only a few on film and television news. Yang Min and Fu Xiaoli (2018) took historical discourse analysis as the theoretical framework, combined with the corpus method, to analyze and study the changes in the US mainstream media's coverage of Hillary Clinton's mail gate, and revealed the discourse strategies of the mainstream media. Lin Yuanbiao and Xu Jiachen (2020) built their own corpora to analyze foreign media's news reports on the 70th anniversary of the National Day from the aspects of headlines, English expressions, etc., and found that China's national image has improved, often linked with prosperity and development. Bai Lifang and Li Pengfei (2021) analyzed the news reports on Wandering Earth by English media based on the corpus method, revealing the focus and evaluation of foreign media on the film.

Most of these studies focus on the news on a certain topic. They compare and analyze the differences between the content reported by Chinese and foreign media or analyze the characteristics of the content reported by a certain media and some of them use the corpus method. This kind of research has developed maturely, but most of it is about political news, rarely involving film and television news, and the research focus is on ideology and other content levels, with little discussion on vocabulary and syntax. The innovation of this paper lies in: 1) Based on the comparable corpus, this paper discusses the differences in vocabulary and syntax between the English translation of Chinese film and television news and the original English news, and puts forward suggestions on the translation of Chinese film and television news. 2) The corpus covers news that introduces TV dramas, films, documentaries and other genres, and the research is comprehensive.

2. Relevant Corpus and Research Tools

Translation language researchers mainly use monolingual translated text corpora or comparable corpora composed of translated original texts. Laviosa (1997) also advocated the use of comparable corpora for translation language research. The corpus in this study is a comparable corpus composed of translated English and original English.

China Daily and Huffington Post are credible media in China and the United States. Therefore, the author chose these two newspapers as the language source of film and television news. The author selected the news about film and television works introduced by China Daily and Huffington Post from 2016 to 2021 as the sample, which covers TV programs, dramas, films, documentaries and other categories, avoiding the prominent features of single category works, making the study more comprehensive and objective. Both are national representative websites, and the selected news topics are consistent and comparable.

The English monolingual comparable corpus involved in the study was built by the author, which is composed of sub database A: English translation of film and television news of China Daily and sub database B: film and television news of Huffington Post. The two sub databases have the same capacity. Sub database A has 217,634 tokens, which are sourced from the Language Tips website of China Daily. The author randomly sampled 441 English translations of film and television news from 2016 to 2021; sub database B has 217,687 tokens, which are sourced from the website of Huffington Post. The author randomly sampled 219 film and television news from 2016 to 2021. All corpora are coded with part of speech through the CLAWS Web Tagger online coding system which adopts CLAWS C5 with a detailed part of speech classification. This study used corpus analysis software WordSmith Tools 7.0 (WST), Antconc and data statistics software Spss 19.0.

3. Results and discussion

3.1 Vocabulary

3.1.1 STTR

The type-token ratio or TTR of the corpus can reflect the variety and richness of words in the text, thus revealing the difficulty of reading. Type refers to the types of all words in the text, while token refers to the total number of all words in the text. In order to reduce the influence of text length on TTR, Scott advocated the use of standard type-token ratio (STTR) to study the degree of vocabulary change, that is, the type-token ratio within a certain number of words. The higher the ratio is, the greater the change of the text vocabulary is, that means richer vocabulary and more difficult reading. The data shown in Table 1 can be obtained by using the WordList function in WordSmith Tools.

Table 1 STTR

	China Daily (Translation)	Huffington Post
token	217,634	217,687
type	14,254	16,159
STTR	45.93%	45.56%

According to the data in Table 1, the STTR (45.93%) of China Daily (Translation) is higher than that of Huffington Post (45.56%), indicating that the word change degree of China Daily is higher than that of Huffington Post, that is, more different words are used in the translations from China Daily which makes vocabulary richer and reading more difficult.

3.1.2 Lexical density

Lexical density refers to the ratio of the content word number to the total word number in a corpus. Content words refer to words with real meaning, containing a large amount of information, including nouns, notional verbs, adjectives, adverbs, etc. On the contrary, functional words do not have complete lexical meaning, but they play a certain role in grammar, including auxiliary verbs, prepositions, pronouns, conjunctions, etc. The lexical density can reflect the information capacity of the corpus and the difficulty of the text. High

lexical density indicates that the text has large information capacity and is difficult to understand. A small lexical density means that the text is less informative, less difficult, and easier to understand. Open the Concord function of WordSmith 7.0, type the code which represents a specific part of speech (for example, AJC represents the comparatives), retrieve the two coded corpora respectively, and finally obtain the following data. Table 2 shows the number and the chi-square test results of content word of each category, the total word number, and lexical density in the two corpora.

Table 2 Word density

	China Daily (Translation)	Huffington Post	X^2	P
Noun	75,255	59,298	2,744.5955	0.000 * * *<0.05+
Notional verb	23,976	25,249	36.7694	0.000 * * *<0.05-
Adjective	19,090	15,271	461.8416	0.000 * * *<0.05+
Adverb	7,966	14,676	2,095.9002	0.000* * * <0.05-
Content word number	126,287	114,494	1,298.9234	0.000 * * *<0.05+
Total word number	217,634	217,687		
Word density	58.03%	52.60%		

The data in Table 2 shows that compared with Huffington Post, P of content word of each category in China Daily translation is 0.000, less than 0.05, indicating that there are significant differences in each value. The two corpora have the most significant differences in the number of nouns and adverbs (X^2 values are 2,744.5955 and 2,095.9002 respectively). The total number of nouns in China Daily (75,255) is much larger than that in Huffington Post (59,298). Compared with modifiers such as adjectives and adverbs, nouns have a greater information load, and the extensive use of nouns will significantly increase the reading difficulty, which shows that the news of China Daily is more difficult to understand than that of Huffington Post. On the whole, the lexical density of China Daily (Translation) is higher than that of Post (58.03%>52.60%), showing a significant difference. It indicates that the news of China Daily (Translation) conveys more information and is more difficult to read.

3.1.3 Average word length

Average word length=the total number of letters/the total number of the words (token) in the corpus. It can reflect the complexity of words. The standard deviation of word length can show the difference between the length of each word in the text and the average word length of the text. The greater the standard deviation is, the greater the difference is. Open WordSmith 7.0 software and run its WordList function to obtain the above two values. Table 3 below shows the statistics of the average word length, standard deviation of word length, and the number and proportion of 6-20 letter words in China Daily translation and Huffington Post.

Table 3 Average word length

	China Daily(Translation)	Huffington Post
Average word length	4.78	4.49
Standard deviation of word length	2.51	2.42
6-20 letter words	68,585	60,384
6-20 Letter word ratio	31.51%	27.74%

It can be clearly seen from the data in Table 3 that the average word length and the standard deviation of word length of China Daily translation are higher than those of Huffington Post (4.78>4.49; 2.51>2.42), with significant differences between the two. In addition, the number of 6 - 20 letter words of the former is significantly higher than that of the latter (68,585>60,384), and the ratio of the former is also higher than that of the latter (31.51%>27.74%). The above data shows that there are many long words in the news of China Daily translation, its difference between word lengths is large, and the use of words is more complex, which increases the reading difficulty. Chen Jiansheng (2010) studied different stylistic texts in Brown Corpus and found that the average word length of academic texts was 4.9, the most formal, followed by newspapers and periodicals, with an average word length of 4.73, and finally novels, with an average word length of 4.31. The average word length of China Daily translation is not much different from that of

newspaper texts in Brown Corpus, which is slightly higher than that of Brown (4.78>4.73), but the average word length of Huffington Post is lower than that of Brown Corpus (4.49<4.73). It indicates that the words used in China Daily translation are more formal than that of Huffington Post, and the degree of formality of Huffington Post news is lower than that of ordinary newspapers and television news, which is more colloquial.

3.2 Syntax

3.2.1 Average sentence length

Average sentence length refers to the average length of sentences in a text, which can reflect the complexity of sentences to a certain extent. In general, if the sentence is longer, it will be more complex to understand it. The standard deviation of sentence length can reflect the difference between the sentence length and the average sentence length in the text. Open the WordSmith software and run the WordList function to get the average sentence length and standard deviation of the two corpora, as shown in the following table:

Table 4 Average sentence length

	China Daily(Translation)	Huffington Post
Number of sentences	9,297	12,647
Average sentence length	23.41	17.21
Standard deviation of sentence length	11.80	11.47

According to the data in Table 4, the average sentence length and standard deviation of sentence length of China Daily translation are greater than those of Huffington Post (23.41>17.21; 11.8>11.47). Leech (1981) proposed that the average sentence length of various styles is 17.8. Daily is much higher than this value, but Post is lower than it. This shows that the film and television news of China Daily translation is more complex and more difficult to read than that of Huffington Post, and the former is more complex and more formal than the ordinary language style, but the latter is simpler and more colloquial than it.

4. Conclusion

Based on comparable corpora, this paper examines the indicators such as the standard type-token ratio, lexical density, average word length and average sentence length of the two corpora of film and TV news from China Daily(Translation) and Huffington Post from the perspective of lexicon and syntax. According to the results of data analysis, the author summarizes the significant linguistic features of the English translation of Chinese film and television news as follows: 1) At the lexical level, compared with Huffington Post, the English translation of China Daily's Chinese film and television news is richer in terms of words, more information loaded and more inclined to use long words, so it is more complicated to understand. In addition, it is more formal than that of ordinary newspapers and magazines, while the Huffington Post's words are colloquial and simple. 2) At the syntactic level, the English translation of China Daily's film and television news tends to use long sentences, which increases the reading difficulty. In order to stimulate readers' interest in China's excellent film and television works more effectively in a short time, the author suggests that domestic translators use more colloquial and simpler English words to translate Chinese film and television news; Avoid using long sentences and reduce the reading difficulty. It is hoped that this paper can give some enlightenment to translators.

References

[1] Laviosa, S. How Comparable Can "Comparable Corpora" Be? [J]. Target, 1997(2): 289-319.

[2] Leech,G. Style in Fiction: A Linguistic Introduction to English Fictional Prose[M]. Beijing:Foreign Language Teaching and Research Press, 1981.

[3] Scott M. The WordSmith Tools[M]. Oxford: OUP, 2004.

[4] Bai LF, Li PF. Corpus based research on English reports of Wandering Earth [J]. Journal of Hainan University, 2021 (1): 71-80.

[5] Chen JS, Cui YN. A corpus based study on the lexical features of the English version of China's Government Work Report [J]. Contemporary Foreign Language Studies, 2010 (6): 39-43.

[6] Hu KB. Introduction to Corpus Translation [M] Shanghai: Shanghai Jiaotong University Press, 2011:28.

[7] Lin YB, Xu JC. A corpus based study on discourse analysis of foreign media English reports on the 70th anniversary of the founding of New China [J]. Foreign Language Teaching Theory and Practice, 2020 (1): 41-49.

[8] Yang M, Fu XL. The Process and Value of "Historical Discourse Analysis" (DHA) Based on Corpus -- Taking the discourse construction of the mainstream media of the United States on Hillary's E-mail Gate as an example [J]. Foreign Languages, 2018 (2): 77-85.