ISSN: 2710-0170

Corpus based study on translator's style --A case study of Peter Pan

Yimei Xiao
University of South China, Heng Yang, China

Abstract

Based on a parrallell-corpus of Peter Pan consisting of the original text, Ren Rongrong's translations and Yang Jingyuan's translations, the study utilizes tools such as Python, CRIE, Emeditor and ABBYY Aligner to compare 17 linguistic structures of the human translations with machine translations. This study reveals that the two human translations have shown different styles. Ren Rongrong's translation is concise in language, full of rhythm, and less difficult to read, while Yang Jingyuan's translation is rich in language and fluent in text.

Keywords

Corpus; Translator's Style; Quantitative Characteristics.

1. Introduction

Since Baker (1993) advocated the application of corpus in translation studies, the research methodology of corpus linguistics has been widely used in the investigation of translator's style, the verification of translation universals, translation training, and translation teaching, among many other aspects, yielding fruitful results (Liu Kanglong & Mu Lei, 2006). With the development of corpus-based translation studies, scholars have begun to focus on translator's style research in recent years, which primarily concerns the unique translational linguistic features of literary translators or translator groups (Huang Libo & Wang Lifei, 2011). Hermans believed that there existed "another voice" in translation besides that of the text's author, namely, the voice of the translator (1996: 27). Baker (2000) defined translator's style as the individuality manifested through a series of linguistic or non-linguistic features that distinguish a translator from others. A translator's style is embodied in the choice of text types and translation strategies, as well as the methods employed by the translator, such as prefaces, postscripts, footnotes, and textual explanations. Zhang Meifang (2002: 57) pointed out that "using corpus-based research to describe, analyze, compare, and interpret elusive and unremarkable linguistic habits can convincingly demonstrate the existence of translator's imprint."

In light of this, grounded in Baker's methodology, this paper attempts to build a small-scale Chinese-English parallel corpus and apply software such as Python, CRIE, Emeditor, and ABBYY Aligner to conduct statistical and quantitative analyses of vocabulary and sentence-level data from Ren Rongrong's and Yang Jingyuan's translations of Peter Pan. The ultimate goal is to explore the translator's styles embodied in these two translations.

2. Corpus Selection and Methodology

The English text of the Chinese-English parallel corpus for Peter Pan adopts the original work created by Scottish novelist and playwright James Matthew Barrie, which was published in the UK and the US in 1911. The Chinese texts are: "Xiaofeixian Peter Pan" (hereinafter referred to as "Ren's Translation") translated by Ren Rongrong and published by Children's Publishing

ISSN: 2710-0170

House, and "Peter Pan" (hereinafter referred to as "Yang's Translation") translated by Yang Jingyuan and published by Nanjing University Press. The word counts of the two translations are 75,812 and 73,844 respectively, which are relatively similar and thus comparable, meeting the criteria for corpus selection.

Han Hongjian (2016), integrating corpus-based translation studies and quantitative linguistics, selected 16 linguistic quantitative features such as word length, sentence length, standardized type-token ratio (STTR), and the proportions of various content words and function words as objects of investigation. Given that Peter Pan belongs to children's literature, the use of various reduplicated words is also an important quantitative feature reflecting stylistic differences. Therefore, this study includes this feature in the investigation. Ultimately, this study identified 17 linguistic quantitative features for analysis. At the lexical level, there are 13 objects of investigation, including standardized type-token ratio, lexical density, high-frequency words, reduplicated words, proportions of nouns, verbs, adjectives, adverbs, numerals, quantifiers, conjunctions, pronouns, and auxiliary words. At the syntactic level, there are 4 objects of investigation, namely, sentence length, proportions of declarative sentences, interrogative sentences, and exclamatory sentences. This paper will analyze the data results from both lexical and syntactic perspectives to provide an objective description of the language use and translation styles of the two translations.

3. Lexical-level Investigation

Professor Mona Baker from the University of Manchester in the UK is one of the earliest scholars to utilize corpora to study translator's style from a lexical application perspective. She examined the translation styles of British translators Peter Bush and Peter Clark, using type-token ratio (TTR) and the usage of "say" and its variations as entry points (Hu Kaibao 2017:15). This study leverages Python statistical tools and the Chinese Readability Index Explorer (CRIE) system to quantify the linguistic features of the texts, enabling the acquisition of basic lexical data from the two translations. Python is employed for word segmentation and part-of-speech tagging to calculate metrics such as TTR, standardized type-token ratio (STTR), word length distribution, lexical density, and high-frequency words. The CRIE system provides counts of nouns, verbs, adjectives, pronouns, and other parts of speech, allowing for the derivation of their respective proportions. Detailed statistics for the two translations are presented in the following tables. The following discussion will explore aspects such as standardized type-token ratio, high-frequency words, word length distribution, lexical density, proportions of content words and function words, and word frequency categories, comparing the linguistic features of the two English translations and subsequently analyzing the stylistic differences between them.

3.1. Type-Token Ratio (TTR) and Standardized Type-Token Ratio (STTR)

Types refer to distinct words in a text, excluding repetitions and ignoring case sensitivity, while tokens refer to all occurrences of words in the text (Baker, 1995). A higher type-token ratio (TTR) in a text indicates a richer and more varied vocabulary used by the author (Baker, 2000:250). The statistics for STTR and lexical density are shown in Table 1.

Table 1: Statistical Analysis

	Ren's	Yang's
Type	58856	57826
Token	6376	7402
TTR	10.83	12.8
STTR	33.09	35.68
Words	75812	73844
Lexical Density	0.8013	0.7957

ISSN: 2710-0170

As shown in Table 1, Ren's translation has the highest number of tokens, followed by Yang's translation, while Yang's translation has the highest number of types, followed by Ren's translation. Differences in the numbers of tokens and types affect the Type-Token Ratio (TTR) and Standardized Type-Token Ratio (STTR). Baker pointed out that the Type-Token Ratio is proportional to the richness and diversity of vocabulary used by the writer. When comparing texts of different lengths, the Type-Token Ratio can be affected by the degree of uniformity in the clustering of types. Therefore, using the standardized Type-Token Ratio is more reliable (Baker 2000:250). As can be seen from the table, Yang's translation has a higher standardized Type-Token Ratio than Ren's translation. This indicates that with the same number of words, Yang's translation employs a more diverse vocabulary. In terms of word count, Ren's translation is the longest, followed by Yang's, suggesting that Ren consciously reduced the number of types during the translation process to make the text easier to read and more accessible for children. From the perspective of translation skopos theory and translator subjectivity, the different linguistic features of the two human translations are actually the result of the translators' different translation purposes and the exercise of their subjectivity. Liu Qiuxi (2013) pointed out that Ren Rongrong has developed personalized translation principles when translating children's literature: fidelity, colloquialism, childlike fun, and contextualization. Yang Jingyuan, on the other hand, employs a range of translation techniques to achieve a translation that is concise, fluent, lively, vivid, and beautiful, resulting in a higher vocabulary richness compared to Ren Rongrong's translation.

3.2. Lexical Density

There are two methods to calculate lexical density: one is to use the Type-Token Ratio (TTR) as lexical density, and the other is to calculate the proportion of content word tokens in the total number of tokens, as proposed by J. Ure (1971) and Michael Stubbs (1986). This study chooses the second method because the corpus size and selection can affect the TTR, which may not accurately reflect the variability of vocabulary. The division of Chinese vocabulary and the distinction between content words and function words remain controversial (Hu Xianyao, 2005). Specifically, this study considers nouns, verbs, adjectives, adverbs, numerals, and classifiers as content words, and the proportion of content words represents lexical density. A lower lexical density indicates relatively reduced information content and reading difficulty (Wang Kefei, 2008). As shown in Table 1 above, Ren's translation has a slightly higher lexical density than Yang's, suggesting that Ren's translation is slightly easier to read.

From the perspective of the proportions of different word classes, as seen in Table 2, Ren's translation has lower proportions of nouns, verbs, and adverbs compared to Yang's, further confirming that Ren's translation is slightly easier to read. Notably, the most significant differences lie in the proportions of adverbs and conjunctions. English has a rich vocabulary of adverbs that are frequently used. In Chinese, adverbs can modify verbs, adjectives, and even entire sentences, functioning as adverbials and complements. Zhang Yisheng (1996) pointed out that a coherent text must contain a certain number of cohesive elements, and the arrangement of sentences and paragraphs should be logical, with intrinsic semantic connections between sentences. In Chinese texts, besides conjunctions, pronouns, and interjections, some adverbs also serve as cohesive elements. In terms of conjunctions, they are crucial in expressing semantic logical relationships within a text, making it a coherent and interconnected whole. Adversative conjunctions are particularly important in achieving logical coherence in argumentative essays (Wang Yang, 2013). English has a larger variety and higher frequency of conjunctions than Chinese. The higher proportion of adverbs and lower proportion of conjunctions in Yang's translation demonstrate its fluency and vivacity.

ISSN: 2710-0170

Table 2: Vocabulary Ratio Data

	Ren's	Yang's
Noun	0.1711	0.1795
Verb	0.2648	0.2758
Adjective	0.0468	0.0467
Adverb	0.0903	0.0975
Numeral	0.0321	0.0302
Quantifier	0.0266	0.0250
Conjunction	0.0250	0.0240
Pronoun	0.1617	0.1400
Auxlliary	0.0894	0.0984

3.3. Reduplicated Words

Reduplicated words, also known as iterative words, refer to words or phrases formed through "syllable reduplication," "morpheme reduplication," "word overlapping," "supra-word overlapping," or "word repetition" (Li Yuming, 2009). They are a unique feature of the Chinese language. Given that Peter Pan is a children's book, the translation style should be imbued with childlike charm. Han Yang (2019) argues that the flexible use of reduplicated words in the Chinese translation of children's literature can enhance the appeal of the original story, improve children's cognitive understanding of Chinese language and culture, and enrich their aesthetic experience of the story. It is thus an indispensable part of the linguistic research on the Chinese translation of children's literature. In this study, Python was used for word segmentation, followed by Emeditor with regular expression input for retrieving the target items. Finally, manual screening was conducted to arrive at the following statistics.

Table 3: Statistics of Reduplicated Words

	Ren's	Yang's
AA	73	50
AAA	2	0
AAB	85	60
ABA	5	8
ABB	32	47
AAAA	1	0
AABB	60	65
AABC	25	32
ABAB	17	8
ABAC	50	51
ABCB	8	6
ABCC	11	15
Total	369	342

From Table 3, it can be concluded that Ren's translation employs the largest number of reduplicated words, followed by Yang's. In terms of quantity, Ren's translation uses more reduplicated words, indicating that it should be more interesting and engaging. This can be further illustrated through the analysis of Example 1 and Example 2:

Example 1:

ISSN: 2710-0170

Original English: One day when she was two years old she was playing in a garden, and she plucked another flower and ran with it to her mother.

Ren's Translation: 她两岁的时候，有一天在花园里玩，摘了一朵花，拿着它噔噔噔跑到妈妈那里。

Yang's Translation: 她两岁的时候，有一天在花园里玩，她摘了一朵花，拿在手里，朝妈妈跑去。

In Example 1, Ren translates "ran" as "dengdengdeng ran," while Yang translates it simply as "ran." The use of "dengdengdeng" vividly portrays Wendy's eagerness and joy in running to her mother with the flower, adding a sense of imagery to the sentence.

Example 2:

Original English: They could hear Nana barking, and John whimpered, “It is because he is chaining her up in the yard,” but Windy was wiser

Ren's Translation: 他们听到南娜在汪汪叫，约翰抽抽嗒嗒地说：“都因为爸爸把它用铁链拴在院子里了”。可是温迪更聪明。

Yang's Translation: 他们听得见娜娜的吠声，约翰呜咽着说：“这都是因为他把她拴在院子里了。”可是温迪知道得更多。

In Example 3, Ren translates "barking" and "whimpered" as "wangwangwang barking" and "chouchoudada" respectively, while Yang translates them as "barking" and "whimpered." The use of "wangwangwang" and "chouchoudada," both being onomatopoeic reduplicated words, brings the images of Nana the dog and John to life, making the reading more enjoyable and melodic.From the analysis of Example 1 and Example 2, it is evident that Ren's translation makes extensive use of reduplicated words, imparting a childlike charm to the translation and enhancing its aesthetic appeal in terms of rhythm, form, and semantics.

4. Examination at the Syntactic Level

This study employed the ABBYY Aligner software to align the two Chinese translations with the original English text. CRIE was used to calculate the number of sentences, word count, and average sentence length, while Python programming was utilized to identify punctuation marks and count the number of words between them, yielding the required data. By examining the average sentence length and sentence features, a comparative analysis of the linguistic characteristics of the four translations at the sentence level can be conducted.

4.1. Average Sentence Length

Similar to the type-token ratio, average sentence length is also a general indicator of a translator's style (Olohan, 2004). In this study, sentence markers were designated as question marks, full stops, exclamation marks, and ellipsis. There is little difference in the average sentence length between the two translations. This indicates that neither Ren Rongrong nor Yang Jingyuan made significant alterations to the sentences themselves, exerting minimal manipulation at the sentence level.

4.2. Sentence Features

The proportions of declarative, interrogative, and exclamatory sentences in the original novel are 0.90, 0.06, and 0.04, respectively. Ren's translation has a higher proportion of declarative sentences than both the original and Yang's translation, a lower number of interrogative sentences compared to Yang's translation, and fewer exclamatory sentences than the original. This suggests that Ren Rongrong flexibly uses punctuation marks, resulting in more declarative sentences and fewer exclamatory sentences than the original. In contrast, Yang's translation has a declarative sentence proportion equal to the original, with slight deviations in the

ISSN: 2710-0170

proportions of interrogative and exclamatory sentences, indicating that Yang Jingyuan still makes appropriate changes to sentence punctuation during translation.

5. Conslusion

This study is based on a one-to-two parallel corpus comprising the English original of Peter Pan and its two Chinese translations. From a quantitative linguistic perspective, statistical tools such as Python, CRIE, and Emeditor were employed to compile data at the lexical and sentential levels of each translation. A comparative analysis was conducted on 17 linguistic quantitative features between the two translations. The aforementioned analysis reveals differences in various aspects, including standardized type-token ratio (STTR), lexical density, proportions of various word categories, high-frequency words, reduplicated words, and sentence length.

Ren's translation exhibits a lower standardized type-token ratio and lexical density compared to Yang's, with the highest word and sentence counts. The frequent use of reduplicated words suggests that Ren Rongrong considered the target audience of children during translation, intentionally reducing the number of types to lower reading difficulty, resulting in a text that is childlike and catchy. In contrast, Yang's translation boasts the highest standardized type-token ratio and lexical density, with the lowest proportion of pronouns among the top ten high-frequency words. This indicates that Yang Jingyuan employs a rich vocabulary, preferring to use names and other expressions to replace frequently occurring pronouns, making the text more in line with Chinese expression habits, fluent, and highly readable.

Due to their unique experiences and backgrounds, human translators often imbue their translations with distinct styles. Although this study, combining corpus technology and quantitative linguistic features, qualitatively and quantitatively reveals translator styles to a certain extent, it has its limitations. The comparative analysis of the two translations is confined to linguistic features. Further research could explore translator styles from non-linguistic perspectives such as context and ideology, thereby broadening and deepening the scope of translation studies.

References

[1] Baker, Mona. Corpus linguistics and translation studies: implications and applications [A]. In M. Baker, et al (eds.). Text and Technology [C]. Amsterdam: Benjamins, 1993: 233-25.

[2] Baker,Mona.Corpora in translation studies: an overview and some suggestions for future reasearch[J].Target,1995(2):223-243.

[3] Baker,Mona.Towards a Methodology for Investigating the Style of a Literary Translator[J]. Target, 2000(2):241-266.

[4] Olohan,Maeve.Introducing Corpora in Translation Studies[M].New York:Routledge,2004.

[5] Peter Pan [M]. Translated by Yang Jingyuan. Nanjing: Nanjing University Press, 2020.

[6] Feng Qinghua. Lexical Translation under Thinking Modes [M]. Shanghai: Shanghai Foreign Language Education Press, 2012.

[7] Han Hongjian, Jiang Yue. A Corpus-based Comparative Study of Linguistic Features in Human and Machine Translations of Literary Works: Taking Three Translations of Pride and Prejudice as Examples [J]. Foreign Language Education, 2016(5): 102-106.

[8] Han Yang. A Corpus-based Study on the Application of Reduplicated Words in Chinese Translations of Children's Literature: Taking Li Wenjun's Translations as Examples [J]. Computer-Assisted Foreign Language Education, 2019(3): 15-21.

[9] Hu Kaibao, Xie Lixin. Corpus-based Translator Style Studies: Connotations and Approaches [J]. Chinese Translators Journal, 2017, 38(2): 12-18+128.

ISSN: 2710-0170

[10] Hu Xianyao. A Corpus-based Study on the Universality of Translation [J]. Journal of PLA University of Foreign Languages, 2005(3).

[11] Huang Libo, Wang Kefei. Corpus Translation Studies: Issues and Progress [J]. Foreign Language Teaching and Research, 2011(6): 911-924.

[12] Li Yuming. A Review of Chinese Reduplication Types [A]. In Wang Guosheng & Xie Xiaoming (Eds.), Issues in Chinese Reduplication [C]. Wuhan: Central China Normal University Press, 2009.

[13] Liu Kanglong, Mu Lei. Corpus Linguistics and Translation Studies [J]. Chinese Translators Journal, 2006(1): 59-64.

[14] Liu Qiuxi. A Study on Ren Rongrong's Translation Thoughts of Children's Literature [D]. Hunan Agricultural University, 2013.

[15] Zhang Meifang. A Corpus-based Survey of Translator's Style: A Review of Baker's New Method [J]. Journal of PLA University of Foreign Languages, 2002(3): 54-57.

[16] Wang Kefei. Corpus-based Translation Studies [M]. Shanghai: Shanghai Jiao Tong University Press, 2012.

[17] Wang Kefei, Hu Xianyao. A Corpus-based Study on Lexical Features of Chinese Translation [J]. Chinese Translators Journal, 2008(6): 16-20.

[18] Wang Yang, Xiao Yi. Differences in the Use of Adversative Conjunctions between English Majors in Normal Universities and Foreign Language Universities: A CEM-based Study [J]. Foreign Language World, 2013(5): 67-75.

[19] The Little Flying Saucer Peter Pan [M]. Translated by Ren Rongrong. Beijing: Beijing Children's Publishing House, 2017.

[20] Zhang Yisheng. The Textual Connective Function of Adverbs [J]. Language Research, 1996(1): 130-140.