A Corpus-Based Comparative Study of Translators' Style: with Seven Versions of Hetang Yuese as Examples

Wei Gong[1 *], Chao Xu[2]

[1, 2]*School of Foreign Languages, Wuhan Business University, Wuhan, P.R. China*
f_author: 851100874@qq.com, s_author: 634193081@qq.com
**Corresponding author: 851100874@qq.com*

Abstract
The popularization of computer and information technology is underscored by the proliferation of corpus, which has been effectively applied to linguistics, lexicography, teaching and translation. Compared with traditional manual methods, the corpus-based translation study can provide more objective and empirical findings in terms of data mining, text retrieval, and statistics analysis. This paper aims to disclose the styles of seven translators through the data mining into the self-constructed corpus of *Hetang Yuese* (known as *Moonlight over Lotus Pond* in English), including WangT, ZhuT, Goldblatt T, YangDT, PollardT, LiT, and XuT. Based on the statistics, the paper finds that the English-native and non-English-native translators share the similarity of simplification and explication, but differ in sentence difficulty and discourse readability. Therefore, the translators are suggested to combine the advantages of both English and Chinese translators, for better transmission of Chinese prose and exchange of world literature.

Keywords: *Corpus Technology, Data Mining, Translators' style, Moonlight over Lotus Pond*

1 INTRODUCTION

With the development of informatization, networking, and digitalization, we have entered the "Internet +" era, so that a variety of advanced computer information technology, multimedia technology and digital technology have been applied to major fields, including the subject of science, engineering, agriculture, medicine, management, and even humanities. Among all the cut-edged technologies, corpus has been proved to be a quite practical and advantageous tool in translation studies by many scholars [1] [3] [5] [9]. Chronologically, Laviosa divides the development of CTS into three periods: the dawn of CTS (1993–1995), the establishment of corpora in translation studies (1996–1999), and the spread of corpora across languages and cultures (2000–) [7]. Corpus-Based Translation Studies has gained great momentum in the past three decades.

Recent years has witnessed the translation of Chinese literature in full swing, but the question of who is supposed to be the subject of the translation is quite controversial. English native translators who are proficient in foreign languages can ensure that the translation is authentic and fluent. However, Chinese translators who may not be as so sensitive to foreign languages as native translators and occasionally render some unidiomatic expressions, can better understand the profound meaning of the original texts. Therefore, who should take the responsibility of Chinese literature translation is a question worth attention and exploration.

Taking Zhu Ziqing's classic prose *Hetang Yuese* (abbreviated as HTYS hereinafter) as an example, this paper analyzes the style of the native translators, Chinese translators of HTYS with the aid of corpus technology under the guidance of translation theories.

Composed by Zhu Ziqing in 1927, HTYS has been included in Chinese textbook for middle school students and remains a famous piece of Chinese lyrical prose. The author took advantage of the tranquility and beauty of the lotus pond over the moonlight to express his dissatisfaction with reality and his yearning for freedom. After the creation of Chinese source text, a number of translated versions have been publicized, with the first one appearing in 1985 and the latest one in 2019, proving the high literary value of HTYS.

B. Fox et al. (Eds.): IC-ICAIE 2022, AHCS 9, pp. 1416-1421, 2023.
https://doi.org/10.2991/978-94-6463-040-4_211

This paper will take seven versions as examples, four created by Chinese translators, two by overseas translators, and one translated under the cooperation of Chinese and foreign translators (see Table 1 for detailed information).

Table 1: Different versions of HTYS

Translator	Text	Title
Zhu Ziqing	ZhuS	*Hetang Yuese*
Wang Jiaosheng	WangT	*Moonlight over the Lotus Pond*
Zu Chunshen	ZhuT	*Moonlight over the Lotus Pond*
Goldblatt	GoldblattT	*The Moonlit Lotus Pond*
Yang Xianyi & Gladys Yang	YangDT	*Moonlight over the Lotus Pond*
David Pollard	PollardT	*The Lotus Pond by Moonlight*
Li Ming	LiT	*Moonlight over the Lotus Pond*
Xu Jingcheng	XuT	*The Moonlit Lotus Pool*

The above table offers an overview of the subject of the study, namely the source text of HTYS and its seven translated version. The first version, produced by Wang Jiaosheng and published in 1985 in the fifth issue of the Journal *the World of English*, was titled *Moonlight over the Lotus Pond* (hereinafter referred to as "WangT" in corpus). The second edition, presented by Zhu Chunshen, was published in the first issue of *Chinese Translators Journal* in 1992 under the same title of *Moonlight over the Lotus Pond* (hereinafter referred to as "ZhuT" in corpus). In 1995, the American sinologist Howard Goldblatt published the English translation *The Moonlit Lotus Pond* (referred to as "GoldblattT" in corpus) in *The Columbia Anthology of Modern Chinese Literature*. In 1999, Yang Xianyi and his wife Gladys Yang co-translated and published *Moonlight over the Lotus Pond* (referred to as "YangDT" in corpus) in the fifth issue of *English Language Learning* Journal. In 2000, David Pollard translated the text titled with *The Lotus Pond by Moonlight* (named "PollardT" in corpus). The sixth translation is *Moonlight over the Lotus Pond* by Chinese translator Li Ming (named as "LiT" Hereinafter) in 2006 in *Translation Criticism and Appreciation*. The seventh latest translation is presented by Chinese translator Xu Jingcheng in 2019 titled *The Moonlit Lotus Pool* (hereinafter referred to as "XuT" in corpus).

After the collection of the research objects, the corpus will be established for study on translators' style based on the data mining of the target corpus, to promote the better transmission of Chinese literature.

2 LITERATURE REVIEW

After the corpus was applied to the field of translation research in the 1990s, people were no longer limited to the perceptual understanding of the translation, but began to use the powerful computing power of the computer to calculate the relevant data of the translation and analyze the characteristics of translation. Previous studies mainly focus on translation teaching, lexicography [5], corpus construction [10], translation universality [2] [8], and translator's style [4]. Baker points out that translation studies has inherited from literary studies its preoccupation with the style of individual creative writers and from linguistics the preoccupation with the style of social groups of language users [3]. Since the 1990s, Corpus-based translation research has always been a hit due to its positivity and objectivity.

Through the further search on previous literature of HTYS, it is found that in terms of the research object they mainly concentrate on Zhu Chunshen and Yang Xianyi's translation, with only occasional references to other version. And in terms of the number of translated versions in previous studies, most of them focus on single version, the vast majority of which center on Zhu Chunshen's translation. Comparative studies are also favoured in academic community, again mostly on the comparative analysis of Zhu Chunshen's translation and Yang Xianyi's translation. There are also some literatures that focus on the comparison of 3 or 4 translations. But quite few study 5, 6, or 7 translated versions due to time and energy limits. And there are relatively few systematic and comprehensive studies on multiple translations. That is what the corpus can deal with efficiently.

3 RESEARCH DESIGN

3.1 Research Corpus

The target texts can be downloaded from the official website or transformed into electronic texts from paper ones via OCR tech, including the source text of Zhu Ziqing's *Hetang Yuese* (abrreviated as HTYS hereinafter), Wang Jiaosheng's translated text (WangT), Zhu Chunshen's TT (ZhuT), Goldblatt's translated version (GoldblattT), Yang Xianyi and Gladys Yang's cooperated render (YangDT), David Pollard's version (Pollard), Li Ming's translation (LiT), and the latest version rendered by Xu Jingcheng (XuT). The author attempts to build a corpus of the source text and 7 translation version to carry out empirical studies.

3.2 Research question

This study attempts to examine the translators' style based on the self-build corpus of the source text and 7 translated version to respond to the following questions:

- What are the similarities of the seven translated versions of HTYS?
- What are the differences between the English native translators and non-native translators of HTYS?
- What are the implications of the translators' style on the future translation of Chinese literature?

3.3 Research Tool

In order to address the above questions, some corpus tech and software will be utilized to replace the time-consuming and energy-draining manual work, including,

- Editplus, for text editing and cleaning;
- Treetagger, for automatic text annotation;
- Antconc 4.0.5, for corpus construction and data mining.

3.4 Research Procedure

After the whole design of the research, the first step is corpus collection from the official website or transformation via OCR, to get the source text and seven translated versions.

The next step is corpus cleaning though the application of Editplus, to check the correctness of the texts, the spelling of words, the deletion of extra blank spaces, lines, and unreadable code.

Then, the cleaning raw texts are loaded to Treetagger for automatic annotation of the part of speech and checked manually to avoid mistakes. The annotation can realize proliferation of the texts, that is, in-depth data mining from multiple perspectives for diversified research purposes.

After the POS tagging, the corpus can be loaded into Antconc, with which such data as Type/token ratio for lexical diversity exploration, Content words/token ration for vocabulary density, Mean sentence length for sentence difficulty, and Flesch Reading Ease score for discourse readability.

Thereafter, the seven translators' style can be exposed with the aid of specific data obtained from corpus technology.

4 DATA ANALYSIS

4.1 Vocabulary Diversity: TTR

In corpus, token means the total number of words appeared in a text while type refers to the how many types of words existing where the same words will be counted as one type. According to Baker, the Type/Token Ratio (TTR) is proportional to the diversity of the writers' vocabulary. The tokens being the same, the more types, the more variety of words. Therefore the 7 translated version of HTYS are loaded into Antconc, to count the TTR as follows:

Table 2: TTR of HTYS Translation

Version	Type	Token	TTR
ZhuS	461	854	54.0%
WangT	358	717	49.9%
ZhuT	529	1101	48.0%
GoldblattT	549	1219	45.0%
YangDT	463	936	49.5%
PollardT	436	950	45.9%
LiT	556	1203	46.2%
XuT	529	1015	52.1%

The table illustrates two "universal" features of translation: explication and simplification. The former can be revealed from the token while the latter can be seen from the TTR of the source text and the target text. In Zhu Ziqing's source text, there are 854 words in total with 461 types, whereas the target texts, except Wang Jiaosheng's Translation, contain more tokens than the ST, from 936 at least to 1219 words at most. This shows that the translators tend to use more words especially functional words to explicate the original text when they translate from Chinese to English. Zhu Ziqing's HTYS is the such a classic prose that is concise in word but profound in meaning. So the translators prefer to put it in more plain English. When the author looks into the detail of the target texts, it is found that the reason why Wang's version, the exception, contains only 717 tokens less than the source text is that he omits two poems in translation to make it more easy more the reader to understand the main idea. Hence it is not difficult for us to catch a glimpse of the second characteristics of simplification. The type/token ratio of the source text is 54%, which is much higher then the seven target ones, being 49.9%, 48%, 45.0%, 49.5%, 45.9%, 46.2% and 52.1% respectively. Among all the translated versions, Xu's version translated in 2019 has the highest TTR, while Goldblatt's and Pollard's texts has relatively low TTR. This displays that the lexical diversity of target texts are lower the Chinese text, but can be improved over time. And contrary to our common sense, the

English native speakers (Goldblatt and Pollard) use less diversified vocabulary than non-native speakers (Wang, Zhu, Xu, Yang D) in translating Chinese prose to make it more easy and simple to read. The lexical density which will be discussed in the following section may shed a light on this phenomenon.

4.2 Lexical Density: CTR

After the discovery of lexical variety, the vocabulary density is researched by counting the ratio of content words (abbreviated as CW in the table) and tokens. If the tokens remain the same and there are more content words, it is believed by the scholars that the text contains more information since it use more content words instead of functional words to convey more meaning. After computing in Antconc, the CTR statistics are shown in the above Table 3:

Table 3: CTR of HTYS Translation

	WangT	ZhuT	GoldblattT	YangDT	PollardT	LiT	XuT
Noun	185	267	297	223	222	279	262
Verb	102	148	174	135	134	168	155
Adj.	56	91	97	88	80	95	98
Adverb	58	93	85	67	69	103	92
Content words	401	599	653	513	505	645	607
Tokens	717	1101	1219	936	950	1203	1015
Lexical density	55.93%	54.41%	53.57%	54.81%	53.16%	53.62%	59.80%

The statistics in the table can further prove the former supposition that English native speakers tend to render simple version in translating HTYS. Among the seven versions, the lexical density of Goldblatt and Pollard remains the lowest at 53.57% and 53.16% respectively. The version completed under the cooperation of Yang Xianyi and his wife Gladys Yang adopts 54.81% content words (CW). To contrast, the lexical density of other four versions translated by non-native speakers is relatively higher, with Wang Jiaosheng's translation being 55.93%, Zhu Chunshen's 54.41%, Li Ming's 53.62% and Xu Jingcheng's 59.80%, which show they are liable to use more content words and less functional words in their translation so as to contain more information and fully express the profound meaning of the original text. Accordingly, when translating Chinese prose, we are supposed to use more idiomatic rather than absurd and intelligible expressions so as to appeal to more target readers, and at the same time use more functional words to render more fluent and coherent versions.

4.3 Sentence Difficulty: M. Sentence Length

sentence length, which is calculated according to the average number words, can be an indicator of the difficulty of the text [6]. The statistics of the seven translations of HTYS is shown in the following Figure:

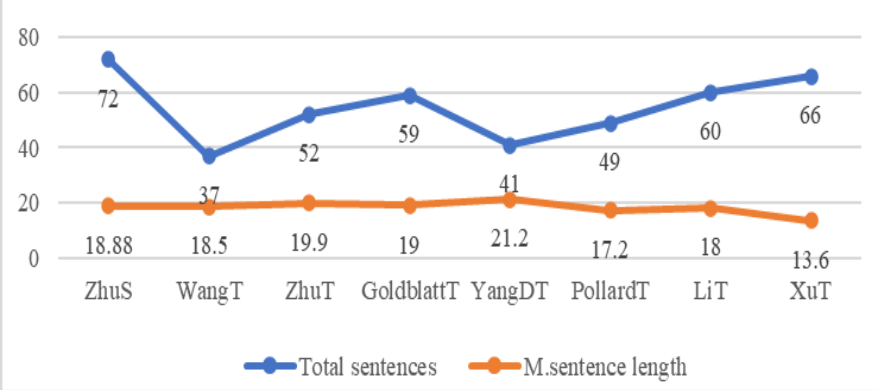

Figure1: Sentence difficulty of HTYS

In this figure, the total sentences of Zhu Ziqing's source text is relatively higher than the translated versions, with 72 sentences in total. This reveals that in Chinese-English translation practices, both the native speaker and non-native speaker translators tend to adopt combination techniques, that is, to combine two or even more short Chinese sentences into longer English sentences using coordinating and subordinating conjunctions "that", "which", "and", etc. The translators' style is the "way of translation" which "distinguishes the translator's work" [9]. As a result, the M. sentence length of three translated texts (with ZhuT being 19.9, GlodblattT 19, and YangDT 21.2) is longer than that of the original one which is 18.88, so as to keep conformity to the English conventions. The rest versions have quite similar M. sentence length, with XuT as an exception, the latest one which has drawn on the previous translations and try utmost to keep fidelity to the ST.

4.4 Discourse Readability: Flesch Reading Ease Score

Readability is one of the factors that reduces the number of readers in the target culture and there is a gap in readability between Chinese translators and English natives in rendering Chinese novels into English. Readability, therefore, is taken as an indicator of translator's style [6]. The seven types of text are evaluated by Readability Analyzer as follows:

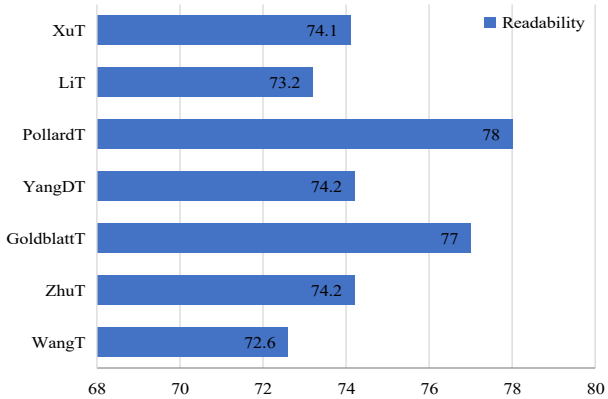

Figure 2: Discourse readability of HTYS

Based on the above bar chart, the readability of two native-speaker translators' texts is much higher, at 78 and 77 respectively. While the texts of non-native speakers hold relatively lower but within the scope of "ease", with WangT at 72.6 which is the lowest, and the other four types (XuT, LiT, ZhuT, and YangDT) roughly similar at 74. According to Flesch Reading Ease score, texts with the reading ease score of 70–79 are fairly easy [6]. Hence all the translated version are not perplex to read and understand, but the non-native speakers translation may also take into account the literariness of the Chinese prose and responsibility of transmitting Chinese culture, which makes the readability of their text correspondingly lower than that of the foreign translators whose main purpose is to popularize the novel or text.

5 CONCLUSION

Within the corpus-based translation studies paradigm, it is aimed by the present research to establish a corpus of HTYS for specific research purposes and to shed a light on the similarities and differences of English native translators and Chinese translators so as to provide reference for future studies in this regard. This study also effectively addresses the original research questions with the aid of corpus. First, Chinese translators do share something in common with overseas translators as for the employment of simplification and explication methods during Chinese-to-English translation in that the type/token ratio of all seven translated texts are lower than that of the source text (the TTR of the source text is 54.0%, while that of the target texts basically lower than 50%). Besides, they also share the similarity in the use of combination technique, which is known from the fact that Zhu Ziqing HTYS contains more sentences in total (72 sentences) and shorter M. sentence length (about 18.88) whereas the translated texts have longer sentences because the translators are prone to use some functional words to combine short Chinese sentences into longer ones in conformity with the habits of target language expression. By further comparison between the data of the seven translated texts, it is found that the English native translators are more likely to render simple and easy-to-read versions because their texts have lower lexical density (53.57% and 53.16% respectively) than that of the texts of Chinese translators (roughly over 54%). Through deeper retrieval into the texts, we found what can account for the lower lexical density may be the more frequent use of functional words (that, which, and, etc.) and less content words. What's more, this

difference can be confirmed by the Flesch Reading Ease score which is over 77 in the native translators' texts but about 74 in the Chinese translators' texts, signifying the former texts are more readable. But whether we should render simple and easy versions when translating Chinese literature definitely is a perplex issue. That is where the future study can dig into by reference to reception theory and with the aid of big data and information technology.

ACKNOWLEDGEMENTS

This research is funded by the Scientific Research Project of Wuhan Business University entitled "A Diachronic Study of Chinese Cultural Image Based on COCA", Grant No. 2020KY002.

REFERENCES

[1] Baker, M. (1999). The Role of Corpora in Investigating the Linguistic Behaviour of Professional Translators. International Journal of Corpus Linguistics, 4(2), 281–298.

[2] Baker, M. (2004). A corpus-based view of similarity and difference in translation. International Journal of Corpus Linguistics, 9(2), 167–193.

[3] Baker, M. (2000). Towards a Methodology for Investigating the Style of a Literary Translator. Target, 12(2), 241-266.

[4] Gang Qiu, Fuxing Su, Le Gao, & Chih-Cheng Chen. (2021). Research on Translation Style in Machine Learning Based on Linguistic Quantitative Characteristics Perception. Sensors & Materials, 33(6, Part 2), 2031–2043.

[5] Hu Kaibao, and Kyung Hye Kim. (2020). *Corpus-Based Translation and Interpreting Studies in Chinese Contexts. Present and Future*. Cham: Springer International Publishing.

[6] Huang Libo. (2015). *Style in Translation: A Corpus-Based Perspective*. Xi'an: Shanghai Jiao Tong University Press.

[7] Laviosa, Sara. (2011). Corpus-based translation studies: Where does it come from? Where is it going? In Corpus-based translation studies: Research and applications , ed. Alet Kruger, Kim Wallmarch, and Jeremy Munday, London: Continuum.13–32 .

[8] Olohan, M., & Baker, M. (2000). Reporting That in Translated English. Evidence for Subconscious Processes of Explicitation? Across Languages and Cultures, 1(2), 141–158.

[9] Saldanha, G. (2011). Translator Style. The Translator, 17(1), 25–50.

[10] Zhang, F. (2021). Application of data storage and information search in english translation corpus. Wireless Networks (10220038), 1–11.