A Corpus-based Study of Styles in Zhu Ziqing's Translated Prose

Xiaomin Yang[1,a,*], Rongjuan Duan[1,b]

[1]College of Foreign Languages, Taiyuan University of Technology, Taiyuan, China

[a] 467048444@qq.com, [b] duan6018312@163.com

*Corresponding author

Abstract: A small corpus based on the English translation of Zhu Ziqing's prose is built, and the corpus analysis software Wordsmith 5.0 is adopted to analyze the English versions by both native translators (Howard Goldblatt) and non-native translators (Zhang Peiji, Xu Yingcai, Zhu Chunshen, Yang Xianyi and Gladys Yang). Their translation styles are analyzed at the lexical level (tokens/types, high-frequency words) and sentence level (mean in words, sentence structure). This study aims to contribute to broader research on translators' styles and provide useful references for future learners engaged in theoretical research and translation practice.

1. Introduction

With China's increasing national strength and cultural influence, there is a strong emphasis on promoting Chinese culture globally. Consistently, there has been a continuous focus on telling a good Chinese story and fostering cultural confidence. Literary translation, particularly in the realm of prose, plays a pivotal role in sharing Chinese culture with the world. Despite its significance, prose translation remains relatively under-explored in genre theory and translation practice when compared to poetry, drama, and novels.

Currently, in Chinese-English translation of prose, there are notable works like *Selected Works of Classical Chinese Prose* by Xu Yingcai, Zhang Peiji's *Selected Modern Chinese Prose Writings*, *Anthology of Chinese and English Prose* by Qiao Ping, and *The Chinese Essay* by British sinologist David Pollard. Additionally, there are some prose pieces translated by Yang Xianyi and Gladys Yang, Zhu Chunshen, Zhang Mengjing, and American sinologist Howard Goldblatt. There are numerous brilliant prose writers, such as Lu Xun, Zhu Ziqing, Yu Dafu, and Bing Xin. Among them, Zhu Ziqing stands out for his distinctive style characterized by "gentleness, elegance, clarity, and profundity". Therefore, this essay selects Zhu Ziqing's famous prose as the object of study.

Different translators carry their own translation styles in the translation process. Previous studies on the translation style mainly focused on a certain translator's style through the analysis of a selected prose piece or a comparative study on the translation style of several translations of one prose piece or a study on the translation style under the guidance of one theory. This essay categorizes translators into two groups: native translators and non-native translators with varying cultural backgrounds, each influenced by their native language. It aims to analyze their translation styles by examining lexical and syntactic features[1]..

2. Corpus-based Translation Studies on Translator's Style

For a long time, translation has only been treated as a reproduction of the meaning and style of the original text, with less emphasis on the creative role of the translator, and their individual style. However, the cultural shift from the 1960s made translation studies shift from the text and the author to the translator, who became visible from invisible, and the study of the translator's style was gradually emphasized.

In the 1950s, the emergence of corpora and corpus linguistics marked a significant development.

Mona Baker creatively introduced corpus-based methods into translation studies. Her framework is based on the idea of stylistic features as translator's fingerprints. Baker (2000:246) believes style in translation involves the translator's choice of the type of material to translate, his or her consistent use of specific strategies-including the use of prefaces or afterwords, footnotes, glossing in the body of the text, and so on. Prominent Chinese translation theorist Professor Liu Miqing (1999) has pointed out that the stylistic features of prose often manifest in vocabulary, syntax, and rhetoric[2]. This essay also tries to explore from lexical and sentence levels[3].

3. Research Design

3.1. Research Corpus

A small corpus is built, including Zhang Peiji's *Selected Modern Chinese Prose Writings*, Book I and Book II, totaling 97 articles, with approximately 183,096 words in both Chinese and English; and selected prose pieces by Zhu Ziqing, including *Green, Spring, Haste, Moonlight over the Lotus Pond, and My Father's Back* (cf. Table 1) in both Chinese and English, totaling about 19,418 words. A few steps are taken to build the corpus, including source text and renderings, text conversion, manual input if necessary, word separation, text alignment, format standardization and separate storage of Chinese and English corpora.

Table 1 English translations of zhu ziqing's prose.

	Zhang Peiji	Zhu Chunshen	Xu Yingcai	Yang&Gladys	Howard Goldblatt	David Pollard
Green			√			
Spring					√	
Hurry	√	√	√		√	
My Father's Back	√		√	√	√	√
Moonlight Over the Lotus Pond		√	√	√	√	

To ensure data balance, this study will primarily focus on the English translations of three selected prose pieces, namely *Hurry*, *Moonlight Over the Lotus Pond*, and *My Father's Back*. The translations by Howard Goldblatt for these three pieces will be referred to as "Goldblatt's," by Xu Yingcai, "Xu's" and the translations of *My Father's Back* by Zhu Chunshen and Zhang Peiji will be collectively referred to as "Zhu's," while the *Hurry* by Yang Xianyi and Gladys Yang, along with Zhang Peiji, "Yang's".

3.2. Research Methodology

This study primarily employs WordSmith (version 5.0) for quantitative analysis and compares different translations at the lexical, sentence and rhetorical levels. Key statistical analyses include Type-Token Ratio (TTR), Standardized Type-Token Ratio (STTR), high-frequency words, average sentence length, sentence structures, repetition, and parallelism. The study aims to explore the translation styles of native and non-native translators through an analysis of these features.

4. Data Analysis and Results

4.1. Lexical Level

4.1.1. Types/Tokens

Types refer to the different lexical words used in a text, reflecting its richness while tokens are the total number of words in a text, indicating the text's length. In contrast to English, which is sentence-oriented with explicit linking, Chinese tend to be short and seemingly loosely connected, relying on inherent semantic connections[4]. Therefore, Chinese-English renderings generally become longer than the source texts.

The Type-Token Ratio (TTR), a measure of the range and diversity of vocabulary, is a crucial indicator for assessing translation characteristics. A big TTR score indicates higher lexical diversity, and thus a richer vocabulary and vice versa.

Table 2 Key features of three prose pieces.

N	Xu's	Yang's	Zhu's	Goldblatt's
tokens(running words) in text	2,771	2,320	2,706	2,894
types(distinct words)	900	805	915	946
type/token ration(TTR)	32.5	34.76	33.81	32.73
mean(in words)	18.05	18.33	18.75	19.41
std.dev.	8.82	9.95	9.35	9.61

Table 2 shows both native and non-native translators have a higher number of tokens, indicating that they have both noticed the differences between Chinese and English during the translation process. Among them, Goldblatt's has the highest tokens (2915) and types (942). The TTR (32.73), suggests a relatively rich and diverse vocabulary usage.

On the other hand, Yang's has the highest TTR (34.76), showing that Yang's translation utilizes a wide range of vocabulary and exhibits a high degree of lexical diversity.

Considering the limited number of different vocabulary and length of a text, however, the Standardized Type-Token Ratio (STTR), rather than TTR, is typically used as a measure of vocabulary diversity for texts or corpora of different sizes.

Table 3 STTR in WordSmith.

Translators	Zhang Peiji	Xu Yingcai	Yang&Gladys	Zhu Chunshen	Howard Goldblatt
standarized TTR	41.50	42.10	40.60	43.10	44.13

According to Huang Libo, the average STTR value for Howard Goldblatt's translations of 17 novels is 44.99, compared to his translations of Zhu Ziqing's prose 44.13[5]. According to Table 3, Goldblatt (44.13) has maintained his consistent style of rich vocabulary usage.

4.1.2. Average Word Length

Words consisting of 2-5 letters are regarded as small or common words. There is a high correlation between average word length and text difficulty. The higher the number of these words is, the more understandable the text becomes. Below is the proportion of these words in the translations by different translators. Below are the results showing the proportion of these words in translations by both native and non-native translators.

Xu: (482+613+565+345)/2782*100%=72%
Yang: (427+486+480+307)/2332*100%=72.9%
Zhu: (503+537+518+341)/2719*100%=69.8%
Goldblatt: (543+590+569+350)/2915*100%=70.4%

According to the results, Yang's is the highest (72.9%), followed by Xu's (72%). This reflects that some non-native translators are reader-centered and prioritize clarity and conciseness during translation.

4.1.3. High-Frequency Words

The analysis of high-frequency words is a fundamental approach in conducting a corpus analysis of literary works. In this essay, the top 15 high-frequency words are listed from 97 selected prose pieces of Zhang Peiji's *Selected Modern Chinese Prose Writings*, as well as 3 prose pieces by Zhu Ziqing, in both their original Chinese and translated versions.

Table 4 Top 15 high-frequency words in 3 prose pieces of Zhu Ziqing and Zhang Peiji's *Selected Modern Chinese Prose Writings*.

N	Word	Freq.	%	N	Word	Freq.	%
1	的	134	6.71	1	的	4,540	7.21
2	我	74	3.71	2	我	1,700	2.70
3	了	62	3.10	3	了	1,424	2.26
4	他	40	2.00	4	是	895	1.42
5	是	32	1.60	5	在	829	1.32
6	着	27	1.35	6	一	823	1.31
7	去	22	1.10	7	他	763	1.21
8	里	22	1.10	8	也	559	0.89
9	在	20	1.00	9	着	511	0.81
10	也	19	0.95	10	不	504	0.80
11	又	19	0.95	11	你	489	0.78
12	一	15	0.75	12	她	451	0.72
13	到	14	0.70	13	有	436	0.69
14	这	14	0.70	14	地	356	0.57
15	有	13	0.65	15	这	329	0.52

According to Table 4, it is evident that the word "的" ranks first in all source prose pieces. Furthermore, the frequent use of "了", "着", and "在" highlights their distinct temporal and colloquial characteristics. The transition from Classical Chinese expressions like "之乎者也" to colloquial ones "的着了吗" reflects the approachable and natural language style prevalent in the prose of that era. Additionally, the frequent use of the first-person pronoun "我", second only to "的", underscores the prevalence of the first-person perspective, and shows a strong lyrical quality of prose. This perspective allows for a subjective recounting of "I" in narrating personal experiences, thoughts, and emotions.

Table 5 Top 15 high-frequency words of 3 prose pieces in target language.

Xu's				Yang's			
N	Word	Freq.	%	N	Word	Freq.	%
1	THE	184	6.61	1	THE	139	5.96
2	AND	84	3.02	2	AND	72	3.09
3	I	82	2.95	3	I	67	2.87
4	TO	80	2.88	4	TO	60	2.57
5	OF	68	2.44	5	A	48	2.06
6	AND	65	2.34	6	OF	43	1.84
7	IN	56	2.01	7	MY	41	1.76
8	HE	35	1.26	8	IN	37	1.59
9	MY	34	1.22	9	HE	35	1.50
10	FOR	29	1.04	10	WAS	31	1.33
11	ME	29	1.04	11	AS	23	0.99
12	BUT	28	1.01	12	ME	21	0.90
13	THAT	24	0.86	13	IT	19	0.81
14	THIS	23	0.83	14	BUT	18	0.77
15	ARE	21	0.75	15	IS	18	0.77
Zhu's				Goldblatt's			
N	Word	Freq.	%	N	Word	Freq.	%
1	THE	165	6.07	1	THE	173	5.93
2	IN	75	2.76	2	OF	92	3.16

3	A	73	2.68	3	AND	88	3.02
4	IN	73	2.68	4	I	86	2.95
5	OF	68	2.50	5	TO	72	2.47
6	TO	67	2.46	6	AND	63	2.16
7	AND	57	2.10	7	MY	46	1.58
8	MY	39	1.43	8	HE	37	1.27
9	HE	34	1.25	9	IN	36	1.23
10	IS	28	1.03	10	THAT	36	1.23
11	ME	27	0.99	11	IT	35	1.20
12	WAS	26	0.96	12	WAS	32	1.10
13	IT	22	0.81	13	ON	29	0.99
14	ON	22	0.81	14	ME	25	0.86
15	BUT	21	0.77	15	HIS	24	0.82

It can be seen from Table 5 that the high-frequency words in Ge's translation are consistent with those in Xu's and Zhu's translations, with "the", "and", I" occupying high positions. In addition, "to", "in", "of", and "a" are also frequently used. According to data, the top five word frequencies in original British and American modern prose texts are: "the", "of", "and", "a" and "to", which shows that native translators consciously adhere to the language norms of the target language and pay special attention to the use of prepositions and articles during translation. Furthermore, in comparison to the second-ranked "of" in the original British and American prose text, Ge's translation exhibits a much higher frequency than that of native translators. This demonstrates that Howard Goldblatt has effectively harnessed the advantages of his native language, resulting in a more formal and literary tone in his translations.

4.2. Sentence Level

Sentence-level analysis primarily involves a comparative study of average sentence length and sentence structure types. Olohan (2004) points out that, in addition to the TTR, average sentence length serves as a general indicator of translator style. It is calculated based on the number of words in a sentence and is typically used as a parameter to measure the complexity of sentences. A higher value indicates a greater level of complexity.

4.2.1. Average Sentence Length

From Table 2, it is evident that the mean in words of Howard Goldblatt's translation exceeds that of native translators. According to Laviosa (1998: 561), the average sentence length of the original narrative text in the English comparable corpus is 24.1 words. Notably, Goldblatt's (19.41) is closer to this benchmark, aligning with the characteristic of long and complex sentences commonly employed in English[6]. As a non-native translator, Goldblatt is subtly influenced by his native English language, reflecting the features of original English prose. However, the value remains below 24.1, suggesting that Goldblatt intends to strike a balance between the source language and the target language. In contrast, the average sentence length of the native translators significantly deviates from 24.1, reflecting their commitment to ensuring clarity and readability for the target readers.

Standard deviation reflects the sentence variety. The higher the standard deviation, the more sentence lengths vary. From Table 2, the standard deviation of the sentences in Yang's and Goldblatt's translations is relatively high, indicating that they employ a combination of long and short sentences in translation to convey the meaning of the original text as well as to ensure readability and acceptability for the readers.

4.2.2. Sentence Structure

Due to structural differences between Chinese and English sentences, translation from Chinese to English frequently entails a process of explicitation or implicitation. According to Seguinot (1988:108), explicitation involves not only making cohesive devices explicit but also conveying

implied meanings not expressed in the source text. In the translation process, the addition of linking words and explanatory information both fall under the category of explicitation[7].

It can be seen from Table 5 that in Goldblatt's translation, the linking word "that" is employed with the highest frequency (36 times). This notable prevalence reflects Goldblatt's consistent use of explicitation strategies. Conversely, in the top 15 word frequencies of Yang's and Zhu's translations, the conjunction "that" is conspicuously absent. The absence suggests that some native translators do not employ a high level of explicitation when using conjunctions.

5. Conclusion

This essay conducts a quantitative analysis of the English translations of three prose pieces by Zhu Ziqing from the lexical, sentence and rhetoric levels through a self-built corpus. Through the comparison and comprehensive analysis of the data, it is found that both native and non-native translators try to faithfully reproduce the meaning and structure of the source text as much as possible in translation. Native translators sometimes use concise language and shorter sentence structures to make their translations more reader-friendly. Besides, the differences between the Chinese and English languages subtly influence the choices of vocabulary and sentence structure made by both native and non-native translators. Non-native translators tend to exhibit more explicitation in sentence structure and , and their lexical choices are closer to the target language.

However, it is essential to acknowledge the limitations of this study. Compared to other literary genres like poetry, drama, and fiction, the number of English translations for the prose works is small. Additionally, due to time limitations, the translations could not be collected completely, resulting in a relatively small corpus. Consequently, the scope of this research remains somewhat narrow, with the analysis primarily focusing on lexical and sentence-level aspects of the text. Future research endeavors should aim to expand the size of the corpus by incorporating translations of a more extensive range of translators' works. Only by delving into a substantial body of material can one discern the "fingerprint" in a translator's style.

References

[1] Liu, Aijun. Comparative Study of Translators' Styles Based on Corpus: English Translations of Zhu Ziqing's Essays as an Example [J]. Technology Enhanced Foreign Language Education, 2020(04):16,101-105.

[2] Baker, Mona. 2000. "Towards a Methodology for Investigating the Style of a Literary Translator?" Target 12(02): 241-266.

[3] Liu Miqing. Contemporary Translation Theory [M]. Beijing: China Translation Corporation, 1999.

[4] Hu, Kaibao. Introduction to Corpus-based Translation Studies [M]. Shanghai: Shanghai Jiao Tong University Press, 2011.

[5] Huang, Libo. Corpus-based Study on Translation Genres [M]. Shanghai: Shanghai Jiao Tong University Press, 2014.

[6] Laviosa S. (1998). Core Patterns of Lexical Use in a Comparative Corpus of English Narrative Prose.Meta, 43(04), 557-570.

[7] Séguinot C. Pragmatics and the explicitation hypothesis[J]. TTR: Tradtion, Terminlogie, Redaction, 1988(02):106-114.