Lecture Notes on Language and Literature (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/langl.2024.070123
ISSN 2523-5869 Vol. 7 Num. 1

A Corpus-based Comparative Study on the Consistency of Writer's Style and Translator's Style

Zhao Xuwei

School of English Language and Culture, Xi'an Fanyi University, Xi'an, Shaanxi, China

Keywords: Translator's style; Corpus; Quantitative stylistics; Lin Yutang; Zhang Zhenyu

Abstract: On the basis of self-built corpora and with the combination of quantitative methods and qualitative analysis, this paper makes a comparative study of the original English works written by Lin Yutang and their Chinese versions translated by Zhang Zhenyu, a well-known translator of the English works of Lin Yutang. Based on the statistics mainly at the linguistic level, including content words ratio, STTR, h-point, entropy and activity, conclusions of this study include: first, the original writer as well as the translator maintains a consistent style during their writing or translating; second, in spite of nuances in certain parameters, the translator's style is highly consistent with that of the original writer; third, it can also be found that quantitative approach, especially quantitative stylistics is applicable and is of higher objectivity for the research of translation study.

1. Introduction

Style is taken as a genre, a variant of the language, or a text type, and Baker, in the first place, saw translator's style as a kind of "thumbprint", which can be explored by his or her regular linguistic and non-linguistic features[1]. From an empirical perspective, the style of a proper and popular target text(TT) is in line with that of the source text(ST), and the translator is expected to try his or her best to "copy" the style of the ST, but to what extent the translator realize that goal remains a difficult field to explore, not to mention the widespread conclusions of translation criticism mostly achieved through qualitative analysis. According to Saldanha[2], Huang[3], translator's style can be studied through two models, source text type(S-type) and target text type(T-type), and T-type, which puts emphasis on the subconscious choices of a translator, is taken as the main model in this study.

2. Quantitative stylistics and its related research

Quantitative stylistics, as a branch of stylistics, focuses on the quantitative statistics of the expression features of works of a certain writer or in certain period, in which corpora and quantitative analysis both play important roles. Leech and Short suggested that stylistic study often entails the use of quantification to back up the judgements which may otherwise appear to be subjective rather than objective[4]. Corpora, as standard representative samples of varieties or languages, provides a particularly valuable basis for an in-depth study of the features of any text. Baker also called for the the use of large computerized corpora in translation studies[5]. Quantitative analysis enables researchers to distinguish genuine reflections of the behaviour of a language from chance occurrences,

thus getting a glimpse of the normality or abnormality of the text or the language through relatively limited statistics, which are much more reliable and generalizable[6].

In China quantitative stylistics study can be divided into two fields: analysis and comparison of writing styles or features; translator identification. In the first field, Huang, et al. successfully applied the achievements of quantitative linguistics to text clustering, language style comparison, laying a theoretical foundation for quantitative study of Chinese works[7][8][9]. Jiang, et al. analysed the quantitative features and differences between human and machine translation of English passive sentences, identifying the translation universals and obvious features of machine translation[9]. Zhan, et al.[10][11][12]made writing style comparisons from different quantitative aspects based on corpora of different works at home and abroad. In the second one, Zhan, et al. utilize corpus-based quantitative research, improving the study of the translator identification by the quantitative features of language structure, thus promoting objectivity and interpretability of translation studies[13][14].

Over years, Chinese translation works are given less attention and as to research of Lin Yutang's original works and their Chinese translations, more efforts should be made. Due to the difficulty in corpora building etc., quantitative research on translational novels in Chinese still plays a small part.

3. Research design

3.1 Research questions

On the basis of previous research, parameters that show the linguistic features of texts are chosen to compare the original English works by Lin Yutang and their Chinese translations by Zhang Zhenyu. Questions to be answered are: Whether there is consistency in the linguistic features and style of original literary works and that of translations respectively; Whether there is consistency in the linguistic features and style between original works and Chinese translations.

3.2 Corpora

As Table 1 shows, this paper is based on the corpus of three original English novels by Lin (hereinafter referred to as Corpus A) and the corpus of corresponding Chinese translation works by Zhang (hereinafter referred to as Corpus B). STs are in general accord with each other in text length so that the influence of text length in the research can be minimized.

Table 1: Research Corpora

Corpora	Titles of the books	Abbreviation	Tokens
Corpus A	A Leaf in the Storm	ALS1	147,216
	Looking Beyond	LB1	128,631
	The Red Peony	RP1	143,844
Corpus B	A Leaf in the Storm	ALS2	125,596
	Looking Beyond	LB2	122,431
	The Red Peony	RP2	149,743

3.3 Choosing of parameters

3.3.1 Content word ratio and STTR

In this study, content words include verb, noun, adjective and adverb both in source language and target language, and the content word ratio is a proper indicator of word richness and variety. Standardized type/token ratio, STTR, is also an important indicator of words richness regardless of

the text's length, so it is of higher reliability when it comes to texts of different length.

3.3.2 H-point, TC

H-point is considered as a critical point or boundary in the rank-frequency distribution of words in certain text. Most of the words before h-point are function words, while most of the words after it are content words, which can be used to measure linguistic type and stylistic features. Thematic concentration (TC) reflects the extent to which the text focuses on a particular topic[15]. Entropy here is calculated based on the probability of occurrence of a word in a text, entropy can also be used to show the richness of words in certain text.

3.3.3 Activity

Liu describes activity as "the ratio of the total number of verbs in the text to the sum of the total number of verbs and adjectives"[16]. Zörnig et al. expressed the interaction between these two part-of-speeches in the following formula[17]:

$$Q = \frac{\mathrm{V}}{\mathrm{V} + \mathrm{A}}$$

where Q is the activity; V is the frequency of verbs and A is the frequency of adjectives. If Q>0.5, the text can be considered as active; If Q<0.5, it can be considered as descriptive and if Q=1, the text is regarded as extremely active[18].

4. Application and analysis

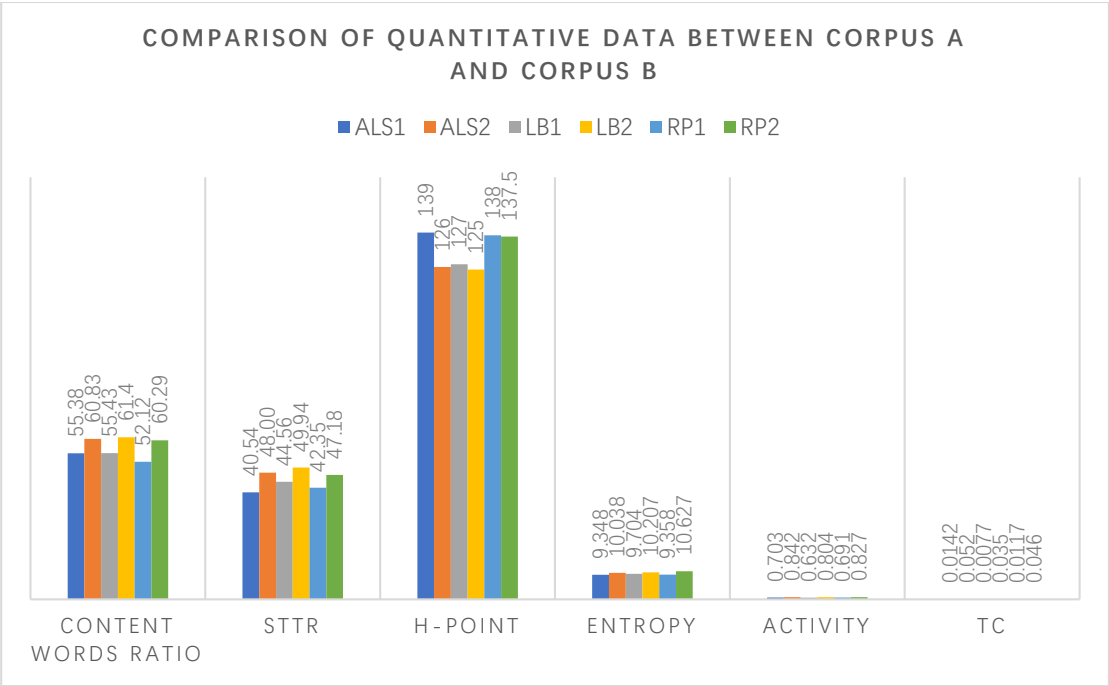

Figure 1: Comparison of Quantitative Data between Corpus A and Corpus B

4.1 Content words ratio and STTR

From Figure 1, the content words ratio as well as STTR of Corpus A is generally lower than that of Corpus B, indicating that the translator tends to use richer and more various words in translations.

The dispersion of content words ratio and STTR of Corpus A is 1.46 and 1.38, showing a high degree of consistency in the original writer's writing habit, while that of Corpus B is 0.37 and 1.04, which indicates an even higher degree of consistency in the translator's choice during translation.

4.2 H-point, TC and entropy

The value of h-point, TC, and entropy in Corpus A and B barely shows high dispersion(Corpus A: 5.1, 0.156, 0.002; Corpus B:5.3, 0.224, 0.006), and both the original writer and the translator maintain a highly consistent writing or translating style. The h-point of Corpus A is higher than that of Corpus B in varying degree, which means that the original writer, in his works uses relatively more function words and the translator chooses more content words in his translations. The value of Corpus B's TC and entropy, accordingly, shows a higher degree than that of Corpus A, and these three values can also illustrate that the translator prefers richer vocabulary and a more explanatory translation for the readability of target readers.

4.3 Activity

The dispersion of the value of activity in Corpus A is 0.03, while that in Corpus B is 0.01, showing little difference in the original writer or the translator's choice in the narrative features of the text. Corpus B has a higher degree of activity, which, as professor Lian summarized, is the result of the favour of verbs in Chinese[19], and based on the previous quantitative research by Xu(2021), the activity of translational Chinese is 0.82, which is close to the result in this study, average activity 0.824.

The original writer and the translator both maintain a consistency in their writing or translating style respectively, and despite nuances in certain parameters as a result of the translator's initiative, the consistency of the linguistic features and style of the original writer and the translator can be examined by SPSS nonparametric test($p=0.317>0.05$), and the result shows consistency without any differences between the statistics obtained. Therefore, the linguistic features and the style of the original writer and the translator share great similarities, and the translator strives to "copy" the original writing style of the ST, which realizes a great reader's response as we can find in the Chinese literary circle or in such comments on Zhang's translations as "fluent and natural"[20].

5. Conclusion

Taking Lin Yutang's English works and Zhang Zhenyu's Chinese translations as an example, this paper tries to utilize quantitative analysis to make a comparison of the style of original writer and translator, and on the basis of statistics, both the writer and translator maintain a highly consistent style in their works and translations. This study serves as convincing evidence for the fact that the writing style of the original writer, Lin Yutang is of high consistency of the translator, Zhang Zhenyu, and for the excellence of Zhan Zhenyu, a translator who dedicated his whole life to translation. But this study mainly focuses on the lexical level, and more statistics on linguistic features are needed to depict a whole picture of the style of a given writer or translator, thus leaving more room for further study in related field.

References

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

[2] Saldanha, G.(2011) Translator style: Methodological considerations [J]. Translator, 17(1): 25- 50.

[3] Huang Libo, Zhu Zhiyu. (2018) Corpus Translation Studies: Research Objects and Methods [J]. Chinese Foreign

Language, 9(6):28-36.
[4] Leech, Geoffrey N. and Michael H. Short. (1981) Style in Fiction: A Linguistic Introduction to English Fictional Prose. Harlow: Longman.
[5] Baker, M. (1993) Corpus linguistics and translation studies: Implications and applications [A]. In M. Baker, G. Francis & E. Tognini-Bonelli (eds). Text and Technology: In Honour of John Sinclair [C]. Amsterdam: John Benjamins, 233-250.
[6] McEnery, Tony & Andrew W. Corpus Linguistics. (2006) An Introduction[M]. Edingburgh: Edingburgh University Pres.
[7] Huang Wei, Liu Haitao. (2009) Application of Quantitative Features of Chinese Corpora in Text Clustering[J]. Computer Engineering and Applications, 45(29):25-33.
[8] Chen Xinying, Li Wenwen, Wang Yan. (2012) The Application of Quantitative Features in Comparison of Language Style and Author Judgement[J]. Computer Engineering and Applications, 48(3):137-139.
[9] Jiang Yue, Zhang Yingxian, Han Jijian. (2016). Comparison of Quantitative Features of English Passive Sentences in Human and Machine Translation[J]. Technology Enhanced Foreign Language Education, (169):46-51.
[10] Zhan Hongwei, Huang Sihong. (2017) Interpretation of Classics in the Era of Big Data: Quantitative Stylistic Analysis of Romeo and Juliet[J]. Foreign Language and Translation, 2:63-68.
[11] Liu Zequan, Wang Mengyao. (2017) Application of Multi-dimensional Analysis in Literary Work's Style—Taking the Old Man and the Sea As An Example[J]. Technology Enhanced Foreign Language Education, (177):61-67.
[12] Shen Wei, Wu Yurong. (2021) Quantitative Stylistic Study of the Novels by Ba Jin and Mao Dun At the Lexical Level [J]. Sichuan University of Arts and Science Journal. 31(4):63-71.
[13] Zhan Juhong, Jiang Yue. (2016) Application of Quantitative Linguistic Properties to Translatorship Recognition[J]. Foreign Language Research, (3):95-101.
[14] Zhong Min, Wang Yang. (2020) Applied Research of Authorship Attribution Based on Computation Stylistics and Multilevel Characteristics [J]. Computer and Digital Engineering, (5):1159-1171.
[15] Kubát,M. ,Matlach,V. & R. Cech. (2014) QUITA-Quantitative Index Text Analyzer[C]. Lüdensheid: RAM.
[16] Liu Haitao. (2017) An Introduction to Quantitative Linguistics[M]. Beijing: The Commercial Press.
[17] Zörnig, P., K. Stachowski, I. Popescu, T. Miyangah, P. Mohanty, E. Kelih, R. Chen & G. Altmann. (2015) Descriptiveness, Activity and Nominality in Formalized Text Sequences[M]. Lüdenscheid: RAM-Verlag.
[18] Xu Zuohao, Jiang Yue. (2021) Activity of Translational Chinese: A Study Based on Online Corpora[J]. Foreign Language Teaching and Research, 53(1):113-123.
[19] Lian Shuneng. (1993) Contrastive study of English and Chinese[M]. Beijing: Higher Education Press.
[20] Zhang Xiuyan. (2017) A Forgotten Translator: Zhang Zhenyu, Translator of the English Works of Lin Yutang[J]. Journal of Jimei University, 20(3):111