Journal of Literature and Art Studies, September 2023, Vol. 13, No. 9, 643-649
doi: 10.17265/2159-5836/2023.09.003

A Corpus-Based Study on the Translators' Styles in the English Versions of *The True Story of Ah-Q*

SUN Huan-ni
University of Shanghai for Science and Technology, Shanghai, China

The True Story of Ah-Q written by Lu Xun has been translated into forty different languages, making great contributions to the external dissemination of Chinese culture. This paper adopts the corpus stylistics method to analyze the translators' styles presented by William A. Lyell and Julia Lovell's translations of Ah-Q. Based on the self-built comparable corpus of the two translations, the paper investigates and analyzes them from the lexical and sentence level. In terms of vocabulary, the translations are statistically analyzed from the standardized type-token ratio, word frequency, word length and vocabulary density; at the sentence level, the article studies the average sentence lengths and sentence complexity. The first major finding is that Lovell's vocabulary is richer than that of Lyell, and Lovell's average word length is longer. Also, the average sentence length of both translations exceeds the original text, and Lovell has the highest average sentence length. Secondly, Lovell prefers to use simple sentences while Lyell prefers complex sentences. Nevertheless, both of them strive to make Lu Xun's works understandable to foreign readers, thus understanding Chinese culture better.

Keywords: The True Story of Ah-Q, corpus stylistics method, translators' styles

Introduction

This thesis attempts to conduct a corpus-based analysis on two English versions of The *True Story of Ah-Q*, with qualitative and quantitative analysis, in order to access the characteristics and difference between two foreign translators and find out the reasons. What's more, by comparing the relations among the two translators' background, native culture, translation strategies and norms, it may shed light for other translators concerned Lu Xun's works, and also enables readers to choose a version right for them to understand. In addition, based on the deeper study and discussion with the help of the self-built corpus, the thesis can provide some suggestions to facilitate the output of Chinese culture to the rest of the world as well as enhance our cultural self-confidence.

Theoretical Foundations and Methodology

As a critical research topic of corpus translation, there are two contrast modes concerning translator style: one is the translator style proposed by Baker (2000), which is the research based on analogy focusing on the

SUN Huan-ni, postgraduate student, Master of Translation and Interpreting, University of Shanghai for Science and Technology.

overall translation style differences among different translations, such as vocabulary diversity and sentence complexity. Another is the study of translator style based on parallel corpus, which pays attention to the regular handling of language phenomena in the same source text by different translators in their respective translations, such as different processing of source text narrative styles.

Therefore, this thesis adopts both the modes above—analog mode and parallel mode. Baker's research method only focuses on the translator's translation text, and the analysis process pays a particular emphasis on the translator's subconscious language behavior (Baker, 2000, pp. 241-266), usually without considering the corresponding source text. Regardless of which mode, the studies complement Baker's methodology from different angles. For this reason, we might as well put aside the strict distinction and expand the scope of research, paying attention to the different treatments of special phenomena in the source text or source language in their translated text.

In addition, this paper adopts a method combining qualitative and quantitative analysis, namely, the quantitative analysis stresses on the words in the form of number. And qualitative analysis aims to identify variables for further study as well as formulate future hypothesis.

The research method is an organic combination of qualitative and quantitative research. Sara Laviosa-Braiwaite points out that Corpus translation research paradigm covers different aspects of translation phenomena, intending to reveal the universal and specific characteristics of translation through the interaction between theoretical construction and hypothesis, various data, new description categories, and rigorous and flexible method (Braithwaite, 1996).

Data Analysis and Results

Discussion at Lexical Level

Vocabulary is usually considered as one of the most significant parameters, which can lead to a more detailed discussion about some features of words and then reflect distinctive translator's style, including such main parameters as type/token ratio, word frequency, word length, and lexical density. Hence, it is easier to know how translators utilize words during the translation process through these parameters and data.

Type/token ratio

Type refers to different words in the corpus, or the form of speech that appears separately for the first time (Laviosa, 1998, pp. 557-570). And token refers to all the forms that appear in the corpus. Type/token ratio refers to the proportion between class character and form character, which can be used to explain the change of lexical application in the corpus. The larger the type/token ratio (short for TTR), the greater the change in the vocabulary used in a specific text, and vice versa.

However, because type/token ratio is easily affected by text length or corpus capacity, Scott (2004) proposed standardized type-token ratio (short for STTR) to measure lexical variability. Table 1 shows the detailed information.

Table 1
TTR and STTR of the Two Corpora

	Lyell's version	Lovell's version	BNC Baby Fiction
Types	3,517	3,212	55,481
Tokens	22,198	16,140	996,002
TTR	16	20	5.57
STTR	42.49	45.32	44.77

The data shown above in Table 1 is calculated with the help of software Wordsmith 4.0. As can be seen from the table, the number of types in Lyell's version and Lovell's version are respectively 3,517 and 3,212, and the number of tokens is 22,198 and 16,140 respectively. According to Vinay & Darbelnet, the token is a significant characteristic of explicitation (Vinay & Darbelnet, 1995). Compared with those of Lovell's, the number of tokens in Lyell's version is much bigger, which means that Lyell's version shows the most obvious language redundancy characteristics and expresses the source information more clearly, making it easier to understand. Also, Nida argues that a good translation is usually longer than the original, because a translator is likely to increase the redundancy of the translation by expressing the implied information of the original text (Nida & Taber, 1969). Therefore, Lyell's translation has more explanations for the source information while Lovell is apt to retain the deep meaning for readers to explore themselves.

Moreover, it is clear that the STTR of Lovell's at 45.32 is the highest, while the Lyell's version is the lowest. It indicates that Lovell's version is more varied and abundant in terminology, while Lyell employs the lowest lexical variability. In addition, the STTR of the reference corpus can be regarded as a standard norm of the word richness of the original English language (Hunston, 2002). We know that the STTR of Lovell's is beyond and close to the data of reference corpus, so the features of vocabulary in Lovell's translation are more in line with those of the original English language. That is why Chinese literature has increasingly entered the mainstream book market in the West and Lu Xun's works have also reached more foreign readers by her vivid translations.

Word length

By comparing the word length, we can find the text with the longest words. In fact, the word length reflects the degree of difficulty, and generally, the longer the mean word length is, the more complicated the text may be. As is known to all, the average word length is 4 letters or so in a common text. The average word length is the ratio of the number of real words to the total number of words in the text. Table 2 below shows the mean word length and word length standard deviation (short for word length std. dev.).

Table 2
TTR and STTR of the Two Corpora

	Lyell's version	Lovell's version	BNC Baby Fiction
Mean word length	4.17	4.28	4.32
Word length std. dev.	2.20	2.40	2.22

Table 2 reveals that Lovell's translation is slightly longer than Lyell's from the respective mean word length at 4.28 and 4.17, in another word, Lovell's translation utilizes harder words and reaches a high level of formality than the other one. What' more, Lovell's version is closer to the reference corpus, which means that it is easier than the native language to understand but more difficult than Lyell's version.

And with the help of software Wordsmith 4.0, statistics about word length std. dev. can be obtained easily and accurately. The word length std. dev in BNC Baby Fiction is slightly larger than Lyell's version and smaller than Lovell's version. This shows that Lyell's is little bit easier and Lovell's is more difficult than native English, and especially Lovell's may be somewhat difficult for foreign readers.

Figure 1 below demonstrates the distribution of different word length in two distinct versions, more clearly and intuitively.

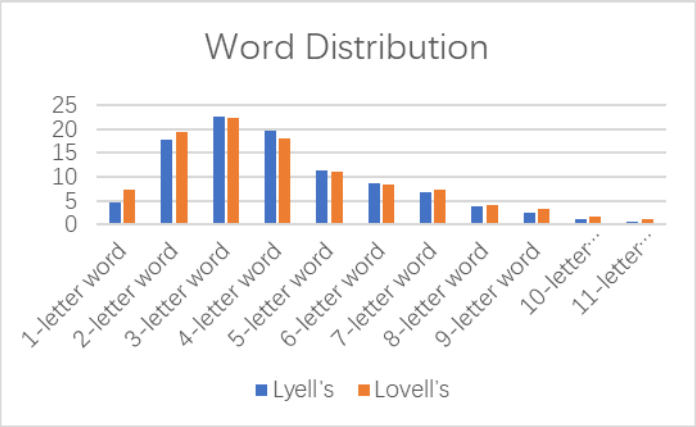

Figure 1. Word length of the two corpora.

In Figure 1, three-letter word ranks the highest among all other words and both the corpora occupy over 20%. The percentage of the other word length from high to low respectively is 4-letter words (around 20%) and 2-letter words (around 18%), 5-letter words (around 11%), 6-letter words (around 8%), and 7-letter words (around 7%). Generally speaking, Figure 1 shows that the average word length of Lyell's version is close to that of Lovell's version.

Usually, a text consists of 1-to-6-letter words, so according to the statistics from Wordsmith 4.0, which lists the frequencies of 1-letter to 6-letter short words and 7-letter to 12-letter long words separately. As is shown in the table, frequencies of short words in the two versions are 84.67 words and 86.52 words respectively, compared with 83.31 words in the reference corpus; while frequencies of long words are 15.02 and 17.95, compared with 16.48 in BNC Baby Fiction. Therefore, it is evident that Lyell uses more complex and longer words in translating *The True Story of Ah-Q*. And it seems that despite that Lyell is not in favor of long words, he tends to narrow the gap with native language by comparing with Lovell.

Word frequency

Word frequency refers to words that appear more frequently than other words in a text, and it shows the explicit vocabulary which a translator is accustomed to utilizing or prefer to choose. Xiao Weiqing indicates that word frequency counting in corpus-based analysis is based on the basic meaning of corpus, which is conducive to the study of vocabulary preference for original or translated texts and the study of translator style (肖维青, 2005, p. 6).

With the aid of Wordsmith, the author compared the first fifteen words in Lyell's version and Lovell's version, and the corresponding sequence in the BNC Baby Fiction.

According to the data collected in Bank of English, the top 5 high-frequency English words are "the", "of", "to", "and", "a" when English is the native language (Olohan, 2004).While the first 5 words are "the", "and", "to", "of", "a" in Translational English Corpus (Butler, 1985).

From the analysis, the top 5 high-frequency words are "the", "and", "to", "are" and "of". In addition, there are altogether 8 words which are shared by three corpora, that is, "the", "to", "of", "and", "a", "was", "in", "it". And three words can be found in the two translations instead of BNC Baby Fiction, for these three words are "Ah", "Q" and "His", which are from the name of major character in the article. By comparing these data, it is

evident that Lyell's version is more similar to BNC Baby Fiction as to the top 5 high-frequency words, which indicates that Lyell prefers to use local vocabulary.

Obviously, "the" ranks first in the three corpora; "to" is listed as the second place in Lyell's version whereas "to" is listed as the third place in Lovell's version and BNC Baby Fiction. Also, most of the high frequency words in the two translations are close to those of native English fiction in terms of word frequency.

Discussion at Syntactic Level

Wang Kefei points out that sentences are important units in translation. In fact, a word is the smallest language unit while a sentence is a language unit consists of words or phrases which can express an integral meaning (王克非, 2004). Hence, it is also necessary to pay attention to characteristics at syntactic level. The pattern of the sentence is determined by the speaker's functional methods, which shows that different speakers always perform different language functions for different purposes.

Butler divides sentence length into three levels: short (1-9 words), medium (10-25 words) and long (more than 25 words). Generally speaking, the longer sentence is that its structure is more difficult and its style is more formal. The average sentence length is the average number of words in a given corpus. What's more, based on the research of English-Chinese parallel corpus, some Chinese scholars also point out that the number of words in the target text usually increases in English translation or Chinese-English translation (冉明志, 2019, pp. 32-37).

Mean sentence length

Average sentence length refers to the average length of translated text sentences. The average sentence length is calculated by period marks, question marks and exclamation points, which is also a general marker of the translator's style. And sentence length standard deviation is the degree of difference between sentence length and average value, which is also related to the translator's style. Table 3 demonstrates the mean sentence length and sentence standard deviations (short for sentence length std. dev) of the reference corpus and two translational corpora.

Table 3
Mean Sentence Length and Sentence length std. dev

	Lyell's version	Lovell's version	BNC Baby Fiction
Mean sentence length	17.61	18.33	12.36
Sentence length std. dev	11.31	12.25	11.70

Table 3 shows that the mean sentence lengths of the two translations are respectively 17.61 and 18.33 respectively, compared with that of BNC Baby Fiction at 12.36. With regard to the sentence length std. dev, the result demonstrates that Lyell's version is lower than Lovell's version at respectively 11.31 and 12.25. Apparently, the mean sentence lengths of the two translations of Lyell and Lovell are longer than that of the BNC Baby Fiction corpus, and Lyell's version has the longest average sentence length. From this aspect, Lyell utilizes the highest complexity in terms of sentence level. Meanwhile, mean sentences std. dev of two versions are close to that of native English fictions.

Sara Laviosa compared the average sentence length of the translated English corpus, and the results demonstrates that the mean sentence length of the former was 24.087, 54% higher than the latter (15.626) (Braithwaite, 1996).This is the reason for the results that two translations are higher than BNC Baby Fiction in

terms of the average sentence length. Besides, the sentence length std. dev. of the Lovell' version is higher than that of the reference corpus, suggesting that Lovell uses the sentences more flexibly.

Complexity of sentences

Zhang Zhenbang divided English sentences into simple sentence, complex sentence, compound sentence and compound-complex sentence. And compound sentence falls into two types: complex sentences and coordinate sentences (胡开宝 & 谢丽欣, 2017, pp. 12-18, 128). Generally, a complex sentence is the mixture of subordinating conjunctions, including "when", "where", "where", "although/though", and appropriate pronouns, such as "that/which", and "who/whom". The following table 4 shows the frequency of subordinators in the two corpora.

Table 4
Frequency of Main Subordinators in the Two Corpora

	Lyell's version	Lovell's version
THAT/WHICH	367	170
WHEN	68	37
WHERE	9	9
WHO/WHOM	62	30
SINCE	26	6
BECAUSE	26	19
ALTHOUGTH/THOUGH	29	39
IF	62	42
Total	649	352
Percentage	2.93%	2.29%

The results from Wordsmith 4.0 demonstrates that there are 1,340 and 910 sentences respectively in Lyell's and Lovell's version. Since that there are 430 sentences more in Lyell's translation than in Lovell's, it employs larger number of main subordinators than Lovell's. On the whole, it is evident that Lovell may utilize fewer complex sentences while Lyell prefers to adopt more complex sentences. We can also find that they have the same frequency in adopting where-clauses. In addition, that/which-clause are the most frequently used in both the corpora, and Lyell prefers to use when-clause while Lovell adopts if-clause more frequently than others.

Conclusion

By applying a corpus-based method to adopt a comparative study of two English versions of the *True Story of Ah-Q*, this thesis discusses particular patterns of different translator's style and demonstrates various characteristics of their translation in a both quantitative and qualitative approach. With the aid of comparable small corpora built by the author, the research is carried out at two major levels, including lexical level and syntactic level. What's more, some possible explanations are also given for the differences and similarity among the corpora. This research investigates the two translator's styles from a descriptive angle and method, and then find what is responsible for their distinctive styles. And this study argues that the translator's native language culture and translation norms might impose an indispensable influence on the translator's style. With regard to their distinctive styles, there are some results listed below:

At the lexical level, Lyell prefers to utilize the shorter words but more difficult words to impart more information while Lovell utilizes a wider range of vocabulary and employs longer words, which is a little bit difficult to read. Also, as to the word frequency, both the corpora is closely in line with the native English expressions, for they try to enable foreign readers to understand Lu Xun's work.

At the syntactic level, the average sentence length of the two translations is much higher than the native English texts. Lyell utilizes more complex sentences while Lovell tends to use relatively simple sentence. However, Lyell's translation is easier to understand than Lovell's.

Nevertheless, this study of translator style is not in-depth enough and not complete enough. And regarding the collection and handling of original texts and statistics are not that precise. Hence, more comprehensive works and more careful efforts are required. What's more, with the explorations and inspirations from this thesis, research concerning translator's style in the near future will be more comprehensive and concrete.

References

Baker, M. (2000). Towards a methodology for investigating the style of a literary translator. *Targets, 12*, 241-266.

Braithwaite, L. (1996). *The English Comparable Corpus (ECC): A resource and a methodology for the empirical study of translation*. UK: University of Manchester.

Butler, C. (1985). *Statistics in linguistics*. UK: Basil Blackwell.

Hunston, S. (2002). *Corpora in applied linguistics*. UK: Cambridge University Press.

Laviosa, S. (1998). Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta: Translators' Journal, 43*, 557-570.

Nida, E. A., & Taber, C. R. (1969). *The theory and practice of translation*. Leiden: Brill Academic Pub.

Olohan, M. (2004). *Introducing corpora in translation studies*. London, New York: Routledge.

Vinay, J. P., & Darbelnet, J. (1995). *Comparative stylistics of French and English: a methodology for translation*. Amsterdam/Phila-delphia: John Benjamins Pub-lishing Company.

胡开宝, 谢丽欣. (2017). 基于语料库的译者风格研究:内涵与路径. *中国翻译, 38*(02), 12-18+128.

冉明志. (2019). 《阿Q正传》三个英译本中文化专有项英译策略对比研究. *四川文理学院学报, 29*(03), 32-37.

王克非. (2004). *双语对应语料库研制与应用*. 北京: 外语教学与研究出版社.

肖维青. (2005). 自建语料库与翻译批评. *外语研究, 4*, 6.