

A Corpus-based Study of the Two Chinese Versions' Styles of *The Conquest of Happiness*

Hou,Ranran

Guangzhou College of Applied Science and Technology, Zhaoqing, Guangdong, 526072, China

Abstract: *The Conquest of Happiness* is a talented philosophical book written by Bertrand Arthur William Russell, a talented philosopher in the 20th century. This paper makes a contrastive study of the translation of *The Conquest of Happiness* to make a meaningful exploration for the spread and popularity of Russell's ideas in China as well as the translators' choice and operation of the translation by adopting a combination of quantitative and qualitative methods to systematically analyze the two Chinese books of *The Conquest of Happiness* from the lexical and syntactic levels.

Keywords: *The Conquest of Happiness*; Fu Lei's version; Huang Han's version; The style study

DOI: 10.62639/sppjess04.20250204

1. Introduction

Mona Baker(1993) predicts that the use of the corpus of the source text and the translated text, with the improvement of the methodology of the corpus, will enable translation researchers to reveal "the essence of the translated text as a medium of communication activities". Many articles have studied it from various angles, but few scholars have studied the differences existing in the two Chinese versions of *The Conquest of Happiness*, and most of the previous studies have focused on qualitative research and failed to analyze them objectively and scientifically. For translation studies, the use of corpus means new ideas and methods. Corpus translatology emphasizes the creative combination of data statistics and theoretical analysis, and the positivism of corpus data also ensures the reliability of translation studies. The two versions(Fu's and Huang's versions) selected are at different times. A good comparative study of the differences between translations can be carried out. Based on this, this paper uses a combination of quantitative and qualitative methods to systematically analyze the two Chinese books of *The Conquest of Happiness* from the lexical and syntactic levels. Through the self-built small corpus and with the help of the retrieval software WordSmith, Antconc, CUC_Paraconc, this paper makes a search and analysis of the two Chinese versions of English philosophical prose *The Conquest of Happiness* translated by Fu Lei and Huang Han. So, while literary works are analyzed by exerting the corpus-based way, English learners' are bound to optimize their learning efficiency and especially bolster their ability of the dynamic mixture of wording and phrasing, structure planning, subjective consciousness, and attitude.

2. Literature Review

The direction of domestic research on Russell's view of happiness is to simply list and analyze one or two points in Russell's view of happiness, such as love, friendliness, and happiness, or to explain Russell's view of happiness from a macro point of view, then contact the individual phenomena in real life to discuss its practical

(Manuscript NO.: JIESS-25-4-62003)

About the Author

Hou,Ranran (1995-), Female, Han Chinese, ShanxiLvliang, Master's degree, Guangzhou College of Applied Science and Technology, Field of study: Translation Theory and Practice, Corpus Translation

guiding significance and enlightening function or combine several major views of happiness of Russell with the views of other philosophies such as Aristotle, Spinoza and others to make a comparative study; a small number of researchers combine Russell's view of happiness with Chinese classics. Zhang Hairen (1992) made clear the close relationship between happiness and desire in Russell's ethical thought. What paves the way for the greatest happiness is the conflict of desires reduced to the lowest. In addition, he specifically pointed out: "Russell discussed in his thought of the supreme moral law: the desire for one's happiness has wreaked havoc in ancient and modern times, but of course, there are still some sages longing for the well-being of all mankind." These people are great moralists who fulfill the highest moral law and are models for people. In her paper, the Ethical details of Love and knowledge-Russell's View of Happiness, Qian Ning (2008) emphasized that Russell's view of happiness focuses on "the typical work of personal ethics and putting individual happiness at the core". In addition, she also specifically pointed out: the purpose and practice of Russell's research on happiness ethics are similar to that of Aristotle, the ancient Greek philosophical giant. However, most of them did not make a detailed and systematic research excavation of the rich connotation of Russell's view of happiness, so they failed to make an objective analysis of it.

3. Corpus-based Translation Research and the Research Methodology

In recent years, regarding the model of corpus linguistics, monolingual comparable corpus, and the parallel corpus have appeared to guide translation activity and improve the quality of translation. The universality of translation refers to the universal characteristics of the translation language which is different from the source language or the target language, including simplification, clarity, and standardization. The research in this field not only proves the existing universal translation hypothesis but also questions it. However, Zhao Qirong and Wang Kefei (2013) found that the universal hypothesis of the translation does not apply to all translation phenomena at all stages. In the study of universal features of translation, manifestation is the most concern by researchers. Huang Libo (2008) examined the manifestation of personal pronoun subjects in English-Chinese translation in both literary and non-literary styles, and Hansen-Schirra et al. (2007) also studied the explicit features in English-German translation. The corpus can provide a large number of real contexts for translation learners and teachers, improve the enthusiasm of scholars and the accuracy of the translation, and greatly expand the horizons and ideas of translation studies.

To draw the objective and scientific research results, quantitative and qualitative methods are inevitably used in this thesis. Firstly, the prime premise of the establishment of parallel corpus is data collection. The two selected versions are the versions of Fu Lei, a modern literary translator, and Huang Han, a PhD in social psychology of Nanjing University. The two translators are in different historical periods in China and are far apart, and there are great differences in language expressions. It has certain research value. At the same time, Russell's original and two translated versions are paper books, and the overall editable electronic text word format is transformed through the scanner and WeChat Mini Program 'Chuantushizi'. However, it is still necessary to manually proofread the paper text word by word to ensure the accuracy of the text. Because the corpus software can only retrieve TXT text, three word texts are converted into TXT format through DOC (X) to TXT batch conversion tool, and they are need to be cleaned and denoised. Secondly, in order to further carry out retrieval and analysis, it is necessary to segment and label the clean corpus text, another important link in the establishment of the corpus. Although CorpusWordParser is a Chinese word segmentation software, it is also very good at English word segmentation. Therefore, this paper uses the corpus tool CorpusWordParser developed by the North Foreign language Corpus to process the three texts. Then, the most critical step in the creation of parallel corpora is to complete the alignment of the English text and the two translations at the sentence level. In this paper, Tmxmall, a powerful alignment tool, is a aid of pairwise alignments between the English text and the two Chinese translations. Lastly,

in corpus retrieval, WordSmith and Antconc are used to describe and count the basic language features of the two versions from the lexical and syntactic levels, and the parameters studied are TTR/STTR, lexical density and average sentence length; the regular expression in Antconc is used to obtain the difference in lexical density between the two translations. At the same time, CUC_Paraconc is used to retrieve the words with significant differences between the two versions, combined with parallel corpus for in-depth analysis.

4. Lexical-level Comparative Analysis

(1) Type token ratio/standardized type/token ratio: word richness

In the area of corpus linguistics, tokens refer to the total number of words in the text, that is, the total number of words in the corpus. Types refer to how many different word forms there are in the text, that is, the number of different words in the corpus, but only the number of type and token can not reflect the essential characteristics of the text. The ratio between them reflects some essential characteristics of the translation text to a greater extent, that is, the variability of words, so type-token ratio (TTR) has become a common method to measure the lexical density of the text (Yang, 2002). However, with the repeated appearance of functional words in the text, the number of text characters is also increasing, so it is inaccurate to use the ratio of class symbols to shape characters as the measurement method of text vocabulary density, so standardized type-token ratio, STTR is born based on of type/token ratio. The so-called standard type/token ratio refers to the calculation of type/token ratio of 1000 words in each text, the average value of several type-token ratios obtained, and the final numerical value is the standardized type/token ratio (Liang, Li & Xu, 2010). The type/token ratio can reflect the differences when the words used by translators: the higher the ratio of type-token, the larger the different vocabularies used in the text; on the contrary, the smaller the different vocabularies. Therefore, the richness and diversity of translators' vocabulary use in different corpora can be compared through the ratio of class symbols to formal characters. However, the difference in text length will directly affect this ratio. In related studies, standardized type-token ratio, generally used to measure the lexical change of corpus text, is more reliable than STTR.

Table 1 The Overview of Standardized Type /Token Ratio

Items	Original text	Fu's version	Huang's version	LCMC
Tokens	57836	52759	47094	854278
Types	5866	6568	6585	42977
TTR	10.14%	12.45%	13.98%	5.03%
STTR	41.52%	50.82%	51.09%	48.32%

From Table 1, the standard type / token ratio of Fu's version is 50.45%, which is lower than that of Huang's version (51.09%), and its word richness is slightly lower than that of Huang's. Although Fu's version is longer than Huang's version, he uses fewer kinds of word types than Huang, so the vocabulary in Huang's version is more diverse than that of Fu's. However, the standard type /token ratio of the two versions is slightly higher than that of the reference corpus LCMC (48.32%), indicating that the words used in the two versions are slightly richer than those in the original Chinese texts, and the vocabulary is varied, thus it can be seen that both translators have good literary literacy and creative talents. Compared with the original work, it is found that the standard type/token ratios of Fu's version and Huang's version are even higher than the original work. This phenomenon shows that all the translations adopt explicit translation strategies, thus increasing the number of characters in the translation.

Then, some typical examples which used the explicitation strategy are extracted from the corpora.

Example1:

S: It is possible to do this process of burying deliberately, and in this way, the unconscious can be led to do a lot of useful work.

Fu: 要有意的去做这番埋藏的手续，是可能的，即在这方式之下，我们可使潜意识做许多有益的工作。

Huang: 我们可以主动进行这种掩藏，这样就可以利用无意识做很多有用的事。

In example 1, both Fu Lei and Huang Han make the English original text explicit, adding the subject “we”, which is in line with the impersonal non-human subject expression commonly used in English, with emphasis on the effect of things on people. In Chinese, personal pronouns are often used as subjects, emphasizing the role or influence of people on things or people, paying attention to subjective thinking, and using people to describe objective things.

Example 2:

S: He wakes up early in the morning while they are still asleep and hurries off to his office.

Fu: 我们来设想一下他的生活看看。清早，全家好梦犹酣的时候，他就得醒来，匆匆的赶到公事房。

Huang: 清晨，家人还在梦乡，他就匆匆起床奔赴办公室。

Pragmatic explicitation is caused by cultural differences, that is to say, translators often have to explain it steadily to overcome the barriers of cultural cognition and equivalence between the source language and the target language. Fu’s version translated “office” into “公事房”, which comes from Lu Xun’s newly compiled story: “So he took the rope, put on the wooden tape, calculated two strings, supported the pole, went to Guan Yinxi’s office to hand in the manuscript, and declared that he was going to leave immediately.” Middle, is a product of the times, is a word that is not common in our daily life. Therefore, Huang uses the word “办公室” is simpler, more straightforward, more colloquial, and close to life.

(2) Lexical density: the density of information

Lexical density is regarded as “a measure of the information density” in a text (Halliday, 2000). Generally, in Chinese, words can be divided into two main categories: content words and function words. Content words are composed of nouns, verbs, adjectives, adverbs. Function words refers to modal words, auxiliary verbs, pronouns, conjunctions as well as prepositions. According to Ure (1971) and Stubbs (1986), lexical density refers to the proportion of content words in the whole corpus. It is proposed that the calculation method is

as follows:

$$\text{Lexical Density} = \frac{\text{Number of Content Words}}{\text{Total Number of Words}} \times 100\%$$

The lexical density of the content words studied in this thesis are shown in the following table according to this calculation method.

Part of speech distribution reflects in part typological features of a language (Qin Hongwu & Wang Kefei, 2009). CorpusWordParser is used to tag the two Chinese versions and with the help of AntConc, the author calculates each number and frequency of each kind of POS in two Chinese versions equally. To have a clear comparison between the two Chinese versions, the author attempts to illustrate differences with a bar graph.

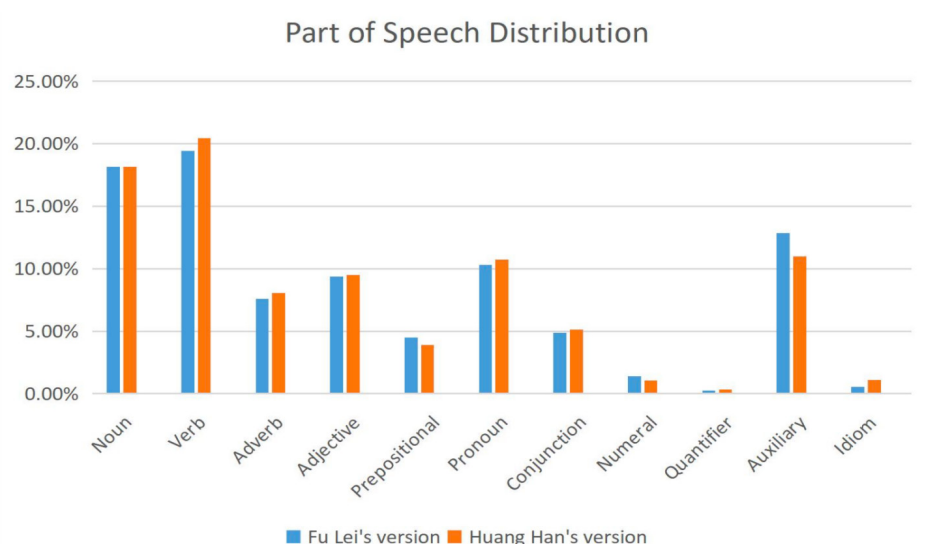


Figure 1 Part of speech distribution

The proportion of part of speech distribution in Fu’s and Huang’s versions are clearly shown Figure1, and they are almost the same, except in the sequence of nouns, verbs, and adverbs that exists some small differences. As can be seen from the chart, the proportion of nouns in Fu’s version stands at 18.16% , while approximately 18.14% in Huang’s version. The frequency of nouns in Fu’s version is slightly higher than that in Huang’s version. Hence, it can be argued that the more frequent use of nouns is the transparent feature of Fu’s version, while the verbs’ frequency in Huang’s version are significantly higher than Fu’s version. So, for readers, they maybe prefer Huang’s version, which is much closer to the original work by Russell.

Table 2 Lexical Density of Content Words

Items	Fu’s version	Huang’s version
Nouns	9582	8544
Verbs	10252	9626
Adjectives	4936	4471
Adverbs	4013	3792
Total	28783	26433
Lexical Density	54.56%	56.13%

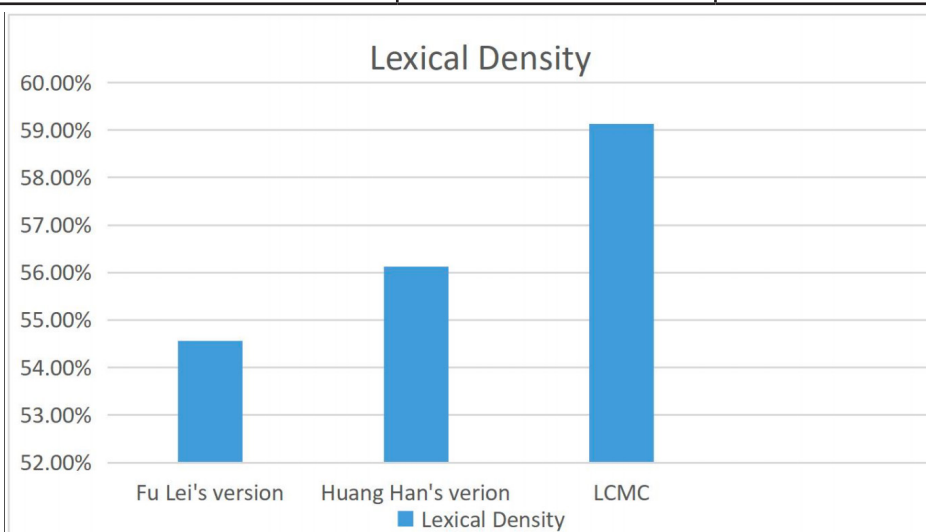


Figure 2 The Lexical Density of the Two Corpora and LCMC

Content words refer to words with stable lexical meaning, including nouns, verbs, adjectives, and adverbs. In the text, notional words usually convey more information than function words, which mainly play the function of connecting grammar and text. The more notional words in the text, the greater the density, the more information is transmitted, and the difficulty increases accordingly. On the contrary, the lower the word density in the text, the less information the sentence carries and the easier it is understood. Thus, lexical density reflects one of the important indicators of the readability and acceptability of the text.

Specifically, the vocabulary density of Fu's and Huang's version is 54.56% and 56.13%, respectively in Table 2. In the two translations, the amount of information contained in the Huang's version is higher than that in the Fu's version, but due to the different times of the two versions, the expression of Huang's version which is lately published is more easily welcomed by the readers. And according to Figure 2, the lexical density of them is lower than that of LCMC. The lexical density means the degree of the text simplification, information capacity and distinguish the translational Chinese from the original Chinese as well. It is known that a significantly high lexical density value is usually around 60% -70% . Thus, both of the two corpora display the simplification strategies for translation style.

Here, some examples are extracted from the two Chinese versions to show their distinctive features in the language.

Example 3:

S: Very few men, I believe, will deliberately choose unhappiness if they see a way of being happy.

Fu: 我相信，倘有一条幸福之路摆在眼前，很少人会胸有成竹地去选择不快乐。

Huang: 我相信，若能看到一条幸福之路，少有人会故意选择不幸。

Example 4:

S: And Browning undoubtedly felt that he was a fine, manly fellow when he denounced Fitzgerald in no measured terms for having dared not to admire Aurora Leigh.

Fu: 当勃朗宁声色俱厉地指斥斐次奇姿特胆敢不赞赏勃朗宁夫人的大作

《奥罗拉·兰格》时，他一定觉得自己是一个出色的，有丈夫气的男子。

Huang: 当勃朗宁有失分寸地公开指责菲茨杰拉德竟敢不欣赏《奥罗拉·利》(Aurora Leigh) 时，他无疑会觉得自己是个有男子气概的好丈夫。

Example 5:

S: A painter, let us say, who has been obscure throughout his youth, is likely to become happier if his talent wins recognition.

Fu: 譬如说，青年时代一向默默无闻的一个画家，一朝受人赏识时，似乎要快乐得多。

Huang: 比如，年轻时不受赏识的画家一旦才华得到认可，其幸福感必然大增。

From the above three examples, it can be seen that Fu uses more idioms unique in Chinese culture. An idiom is not only an important part of Chinese vocabularies but a major feature of Chinese culture, a large part of which is inherited from ancient times, almost every idiom represents a historical allusion. Idioms are concise, incisive, and rich in connotation, and their connotations are often hidden in the literal meaning, and they have a strong sense of rhythm and rich associative meaning, so they are deeply loved by people. For example, "胸有成竹", "声色俱厉" and "默默无闻" reflect Mr. Fu Lei's profound cultural heritage and literary accomplishment, which can also enable readers to receive better cultural edification. By contrast, Huang's translation of "故意", "有失分寸" and "不受赏识" is relatively simple and slightly inferior.

(3) Word frequency: word style

“Word frequency refers to the number of times each word form appears in the text, and word frequency statistics is a basic statistical method in corpus analysis.”(Xiao, 2005) Word frequency statistics is a basic statistical tool in corpus analysis, and the use of word frequency statistics in translation studies is also beneficial to the study of the word style of the source text, the word style of the target text, and even the style of the translation(Zhao, 2018).

Here, somethings valuable are shown, including a word list stored by frequency in descending order and the top ten lexical items in the two versions and the LCMC. The comparisons are displayed in the following table.

Table 3 Top Ten Lexical Items Demonstrating in the Two Chinese Versions and LCMC

Rank	Fu's version			Huang's version			LCMC		
	word	Freq.	%	word	Freq.	%	word	Freq.	%
1	的	5335	10.11	的	4071	8.64	的	51324	6.75
2	在	937	1.78	是	725	1.54	了	13124	1.73
3	是	842	1.6	人	631	1.34	是	12599	1.66
4	他	711	1.35	在	603	1.28	在	10527	1.38
5	人	590	1.12	他	547	1.16	一	10115	1.33
6	而	444	0.84	了	472	1.00	不	7766	1.02
7	不	429	0.81	而	429	0.91	和	7370	0.97
8	了	418	0.79	会	408	0.87	他	5899	0.78
9	一	405	0.77	我	365	0.78	我	5686	0.75
10	我	401	0.76	自己	347	0.74	有	5642	0.74

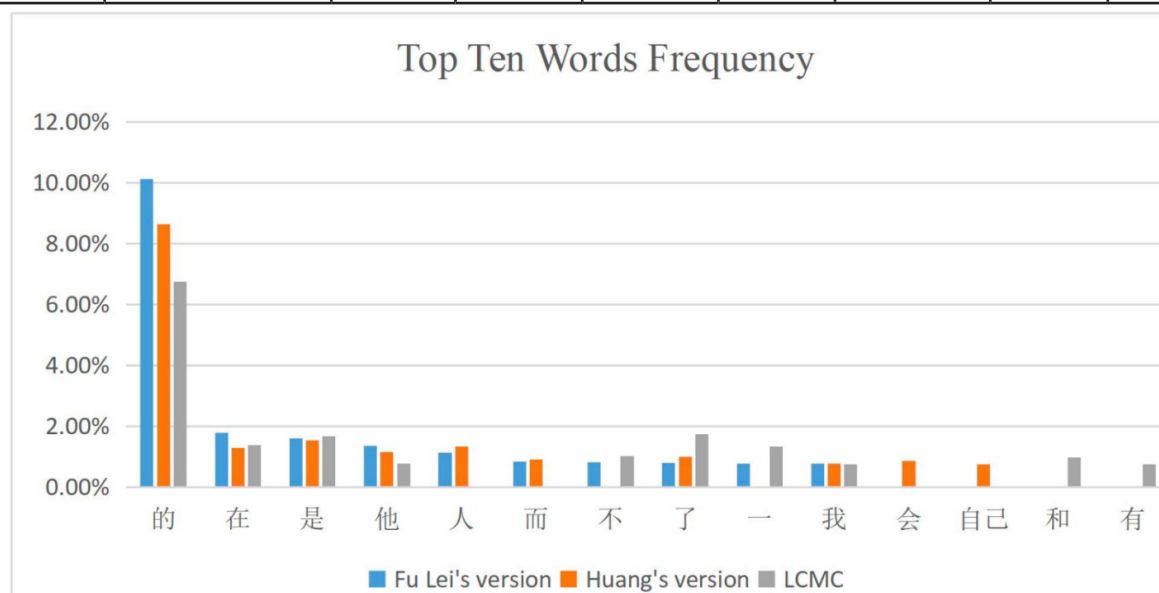


Figure3 A Bar Graph of Top Ten Lexical Items Demonstrating in the Two Chinese Versions and LCMC

For the convenience of demonstrating the difference in the top ten lexical words between the two Chinese versions and the LCMC, the author creates a bar graph to display it. It can be found through the word list that Chinese corpora and LCMC include the Chinese characters like“的、在、是、他、人、而、不、了、一、我、会、和、自己、有”. Through the careful observation of the frequency list of the two versions and LCMC in Figure 3, it is obvious that some other lexical items in the two versions such as“人、而、会、自己”are not found in LCMC. Then, it could be seen from Table 3 that although the order and frequency are somewhat different, the two translations

are very similar to the high-frequency words in the Chinese reference corpus. These high-frequency words are mainly pronouns and function words, such as pronouns such as “我”, “他” and “自己”; and function words such as “而” and “了”, which should have a lot to do with the basic characteristics of the Chinese language and conform to the habits of the Chinese language. In these two translations and the high-frequency words of the reference corpus, the word “de” ranks first, mainly because “de” is one of the most common function words in Chinese, and adjectives in Chinese have to be expressed with the help of “de” when modifying nouns.

That shows at the lexical level, on the one hand, the translator’s unique work will be swingingly influenced by the style of the source text, which reflects the style of the original work on some levels; on the other hand, it will inevitably reflect the linguistic characteristics of the target language itself. In addition, from the view of the high-frequency words in the translations, the word “he” is relatively high. Pronouns are words used to replace or refer to things, actions, shapes, and quantities, and have the syntactic functions of nouns and adjectives in sentences. A personal pronoun is a kind of pronoun, which is used to refer to people or things, such as you, me, him, etc.(Liu, 2007). The two translations are higher than the Chinese reference corpus, which has a lot to do with the stylistic features of the original work. Because *The Conquest of Happiness* is philosophical prose about the philosophy of life, and most of the objects described in the article are described in the third person, the probability of the pronoun “he” will be very high.

5. Syntactic-level Comparative Analysis

Syntactic feature analysis is also an important part of examining the linguistic features of specific languages to translated texts. The syntactic features of specific languages to translation languages can be understood from two aspects: the overall syntactic features and the application features of typical sentence structures. The general syntactic features include the average sentence length of the translated text, the average sentence length, the structural capacity, the frequency of simple sentences and compound sentences, and so on(Hu Kaibao, 2011). The analysis of syntax in translation studies has long been neglected and confined to the rendering skills like how to translate a passive voice across different languages. With the help of Corpus Translation Studies, the research on the syntactic level is largely expanded.

Average sentence length: complexity in text

The average sentence length is expressed by the number of words in the text. The formula is average sentence length = total number of characters / total number of sentences, which is often used to measure the difficulty of text sentences.

Table 4 Average Sentences Length in the Two Chinese Versions and LCMC

Item	Original Text	Fu	Huang	LCMC
Tokens	57836	52759	47094	854278
Sentences	2186	2229	2160	41492
Average Sentences Length	26.45	23.67	21.80	19.98
Std.dev.	13.82	12.08	10.57	14.34

According to Table 4, the items of the average sentence length in three texts and LCMC are shown. And both of the two Chinese versions tend to use more sentences than the source text. The length of average sentence in the source text is 26.45. Then, the average sentence length in the two Chinese versions is 23.67 and 21.80. It could make a conclusion that Fu’s work is more complex than that of Huang’s in view that the average sentence length equals the complexity of the sentence. Compared with Russell’s original work, the average sentence lengths of Fu’s and Huang’s version are apparently much shorter. One English word may be expressed the

corresponding meaning of several Chinese characters, so a Chinese version's text length is usually longer than that of the original English text. As Laviosa(1998: 5) said that in comparison with non-translated language, translated language has a significantly greater mean sentence length. As a paratactic language, Chinese texts usually presents a succession of short and syntactically simple phrases and sentences. In Table 4, it is easy to observe that the extension of sentence length in two Chinese versions is more obvious than LCMC(19.98).

6. Conclusion

Based on the self-built English-Chinese bilingual parallel corpus of *The Conquest of Happiness*, this thesis uses a combination of quantitative and qualitative methods to explore the language features and translation strategies of *The Conquest of Happiness*. From the perspective of language features, in terms of vocabulary, the vocabulary richness of Huang's version is slightly higher than that of Fu's version, and accordingly, the difficulty of reading is slightly higher than that of Fu's version, but there is little difference between them in terms of the amount of information they carry, and they are the same close to the original Chinese text. In addition, the words with high frequency in the two versions are the same. At the syntactic level, the average sentence length of Huang's version is lower than that of Fu's version, but the stability of sentence length is higher than that of Fu's version, which brings readers a better sense of experience. Ultimately, the two versions play an important role for readers and are worth cherishing and studying.

References

- [1] Baker, M. 1993. Corpus linguistics and translation studies: Implications and applications [A]. In M. Baker, G. Francis & E. Tognini—Bonelli (ed.). *Text and Technology: In Honour of John Sinclair* [C]. Amsterdam & Philadelphia: John Benjamins, 233-250.
- [2] Zhang, Hairen, and Wang, Xiaoxi. (1992). *Western Ethicists Dictionary*[M]. China Radio and Television Publishing House.
- [3] Qian, Ning. (2008). *The Ethical Foundation of Love and Knowledge: An Analysis of Russell's Conception of Happiness*[J]. Lanzhou Academic Journal, (7), 13.
- [4] Zhao, Qiurong, and Wang, Kefei. (2013). *The Stage-based Characteristics of English-to-Chinese Translation Language: An Investigation Based on a Diachronic Comparable Corpus*[J]. Chinese Translators Journal, (3), 15–19.
- [5] Huang, Libo. (2008). *Explicitation of Pronoun Subjects in English-Chinese Translation: A Corpus-Based Investigation*[J]. Foreign Language Teaching and Research, 40(6), 454–460.
- [6] Hansen-Schirra S. 2007. *Cohesive Explicitness and Explicitation in an English-German Translation Corpus* [J]. Languages in Contrast: 241-265.
- [7] Yang, Hui Zhong. (2002). *An Introduction to Corpus Linguistics*[M]. Shanghai: Shanghai Foreign Language Education Press.
- [8] Liang, Maocheng, Li, Wenwen, and Xu, Jiajin. (2010). *Corpus Application Tutorial*[M]. Beijing: Foreign Language Teaching and Research Press.
- [9] Fu, Lei. (2019). *The Road to Happiness*[M]. Beijing: Taihai Publishing House.
- [10] Bertrand Arthur William Russell. (2021). *The Conquest of Happiness*[M]. Translated by Huang, Huan. Tianjin: Tianjin People's Publishing House.
- [11] Ure, J. 1971. *Lexical Density and Register Differentiation*. In G. E. Perren & J. L. M. Trim(eds). *Application of Linguistics: Selected Papers of the Second International Congress of Applied Linguistics*. Cambridge: CUP.
- [12] Qin, Hongwu, and Wang, Kefei. (2004). *Corpus-based Analysis of Translated Language: A Case Study on the Chinese Corresponding Structures of 'so...that'*[J]. Foreign Language Teaching and Research, 27(1), 40–48.
- [13] Xiao, Weiqing. (2005). *Building a Self-built Corpus and Translation Criticism*[J]. Foreign Language Research, (4), 60–65.
- [14] Zhao, Tingting. (2018). *A Study on Translator's Style in the Three Chinese versions of Reason and Sensibility Based on Statistics and Analysis of Self-built Corpus*[J]. Modern Linguistics, 6(4), 581–587.
- [15] Liu, Miqing. (2007). *Style and Translation*[M]. Beijing: China Translation & Publishing Corporation.
- [16] Hu, Kaibao. (2011). *Corpus-based Translation Studies: An Introduction*[M]. Shanghai: Shanghai Jiao Tong University Press.
- [17] Laviosa, S. 2002. *Corpus-Based Translation Studies: Theory, Findings and Applications*[M]. Amsterdam & New York: Rodopi.