A Corpus-based Study on Translation Norms of *The Scarlet Letter*

Wu Linyun

1Taizhou Polytechnic College，Taizhou city, Jiangsu Province, China, 225300

Abstract. *The Scarlet Letter* is one of the masterpieces of the famous American writer Hawthorne, ever since its publish, there have been 24 Chinese versions. This paper establishes a monolingual corpus of *The Scarlet Letter*, and classifies the translations into three main categories (1917-1949, 1949-1979 and 1980-now) according to the division of contemporary translation history. By using corpus linguistic methods, this thesis investigates operational norms from lexical and syntactic level, and describes translation norms of all translations of each historical period, therefore reveal how do norms change diachronically.

Introduction

So far, there are 24 different Chinese versions of *The Scarlet Letter*. From the 1930s to 1950s, Zhang Menglin, Fu Donghua, Yang Qirui and Han Shiheng translated the novel. With the development of Chinese economic reform and the open up in the culture communication after the 1980s, there surfs a trend of translating foreign classic works. Consequently, of course, the retranslations of *The Scarlet Letter* come into being. This paper adopts corpus-based approaches to make a diachronic description of the translation norms of all the Chinese versions of *The Scarlet Letter.* It classifies the translations into three main groups according to the division of contemporary translation history, aiming to describe the features and diachronic changes of operational norms.

The research on translation norms is a prevailing topic ever since 1980s. This paper intends to describe translation norms of *The Scarlet Letter*. Thus, the research belongs to the branch of descriptive translation studies. Based on the methodology of corpus research, the author built a DIY corpus of all versions of *The Scarlet Letter* in order to make a diachronic description of the translation norms of the novel.

To be specific, the research questions consist of: What are the translation norms of the different translations of *The Scarlet Letter* in different historical periods? How do the translation norms of *The Scarlet Letter* change diachronically?

Methodology

Graeme Kennedy (1998) elaborates the basic three problems when building a corpus in *An Introduction to Corpus Linguistic,* say, the type of the corpus, the representativeness and balance of the corpus as well as the size of the corpus. The process of corpus normally includes collection, computerization, marking-up and annotation of collected data. The construction of a corpus, whether a large one or a small one, follows a similar routine like this.

The Type of the Corpus—Monolingual. Sara Laviosa (2007) from University of Bari, Italy, made a lecture "Similarity and difference in corpus-based translation studies" in *Conference and Workshop on Corpora and Translation Studies* in Shanghai Jiaotong University in 30 March to 2nd April. In the lecture, Sara Laviosa elaborates on a Research model and methodology of corpus-based translation studies from the following five aspects: "comparative research model, bilingual comparable corpus, bilingual parallel corpus, monolingual comparable corpus, information about the

extratextual factors of the communicative situation (or translating event)". And the researchers DIY corpus is a monolingual one of all the Chinese versions of the novel—*The Scarlet Letter*. The premise of the construction of it is that the language of the translated versions is different from that of the source language and the target language, which is thought as "The Third Code".

Computerizations of the Versions. After getting the raw material from the Internet, bookstores, libraries etc, the author scan the simplified Chinese versions. All the versions are saved in ".txt" (ASCII-American Standard Code for Information Interchange) format for future use. And the total size of the versions is about 3,000,000 characters.

Classification of all the Versions. Zhang Yan (2006) divides the fiction translation from the late Qing into four periods, say, from 1898 - May 4th Period, Post- "May 4th" - 1949, 1949 - the end of 1970s, 1980s - now. It is easy to note that the versions of *The Scarlet Letter* lay in different historical periods. In the 1930s, there are two versions, the first is in 1934 which is later than the first translation of other famous works. In the 1940s, because of the Anti-Japanese War, the unstable society reduces the production of literature works. There are three versions listed between 1917 and 1949, which are all in traditional Chinese. After 1949 and the separation of Taiwan from the mainland, the norms of language began to change gradually to simplified Chinese. In 1954, there boomed up the reformation and simplification of Chinese characters. However, the version in 1954 by Shi Heng is still in traditional Chinese. During the ten years' Cultural Revolution, there is no translation, and the phenomenon is more or less the same of the translation of other European works. From 1980 till now, because of the reform and opening up, it comes to a booming period, many versions emerges.

Corpus Data Analysis Tools. The analysis of corpus data mainly uses the index software Ictclas1.0, Antconc3.2.1w and WordSmith 4.0.

Data Presenting and Discussion

Lexical Features. Lexical density of a text measures the proportion of the content (lexical) words over the total words. There are two ways to calculate the lexical density. One is put forward by Ure(1971)& Stubbs(1986) which is adopted in this thesis. The formula is as follows: Lexical Density=(Number of content words/Total number of words) x 100%. The other is the value of Type/Token Ratio. The formula is: Lexical Density=(Number of different words/Total number of words) x 100%(Yang Huizhong, 2002, p.168). Texts with a lower density are more easily understood and acceptable to the target reader. In addition, lexically dense text has a lexical density of around 60-70%, and those which are not dense have a lower lexical density of about 40-50%.

This research adopts the first way to calculate lexical density. So it is necessary first for us to distinguish content words from other words. In general, words can be divided into two categories: content words and function words. Content words are open class words, while function words are closed class words. Function words are usually grammatical words, which have little lexical meaning. On the contrary, words which are not function words are called content words. This research chooses verbs, nouns, adjectives and numerals as content words because these four classes have relatively stable lexical meanings.After the division of the Chinese versions and POS tagging by Ictclas, the research get the value of lexical density as follows:

Table 1: Lexical Density (%)

	Verbs	Nouns	Adjectives	Numerals	LexicalDensity
Zhang 1934	19.88	18.48	4.88	4.25	46.88
Fu 1937	19.55	16.72	5.00	3.55	44.83
Yang 1942	19.85	17.04	**4.79**	3.39	**45.08**
Shi 1954	19.48	17.09	5.35	3.46	45.39
Hu 1991	19.62	17.45	5.21	3.32	45.61
Zhou 1991	19.94	18.58	6.13	3.98	47.86
LiDong 1996	20.53	17.94	5.39	3.36	47.22
LiYanbo1996	20.71	18.72	5.69	3.41	48.53
Yao 1998	19.43	12.43	9.89	3.37	45.14
Feng 1999	18.49	17.61	5.44	3.13	44.67
Fang 1999	20.93	13.41	**10.68**	3.27	**48.31**
Xie 2000	20.37	17.68	5.39	3.39	45.19
Xiong 2000	20.18	**11.84**	5.54	5.54	**40.91**
Huang 2002	18.84	17.58	5.41	3.19	45.02
Dong 2002	20.01	17.64	5.34	3.33	46.34
Wang 2002	20.95	18.58	5.77	3.42	48.73
Zhou 2002	20.13	17.20	5.05	3.52	45.92
Bai 2003	20.29	18.37	5.24	3.29	47.18
Liu 2003	20.72	18.29	5.55	3.36	47.91
He 2004	19.97	17.79	5.42	3.35	46.53
Jia 2004	20.17	18.29	5.27	3.30	47.02
Wang 2006	20.96	19.17	5.78	3.33	49.24
Zhao 2007	21.77	**18.99**	5.66	4.28	**50.69**
Su 2007	19.34	16.85	4.94	3.41	44.54

From the above chart, it is easy to get that the lexical density of *The Scarlet Letter* is around 45%. Thus, it is not a dense text. The message conveyed in the texts is in a medium level.

The table shows that the percentage of verbs and numerals are relatively static while the percentage of nouns and adjectives vary from time to time. The use of nouns in Xiong Yupeng & Yao Naiqiang's version reaches the lowest level while in Zhao Xunian's version, it reaches the highest level. And an obvious change of adjectives views in the version of Fang Wenhua, which reaches the highest level, while Yang Qirui's version reaches the lowest level. The extremes of the figures will affect the lexical density of that particular translation, and further affect the overall standard of the whole period. They are individual cases. So the lexical density of Yang Qirui (45.08%), Fang Wenhua (48.31%), Xiong Yupeng & Yao Naiqiang (40.91%) and Zhao Xunian (50.69%) are set aside. To be more specific, the lexical density in the first phase, say, 1917 to 1949, is 45.86%, from 1949 to 1979, the average percentage is 45.39%, which is a little bit lower than the overall standard of the first one. Then after 1980s, the average percentages reach 46.63%. And among them, different versions' lexical density varies. The variation can be seen much more clearly and vivid form the following chart.

Lexical Density

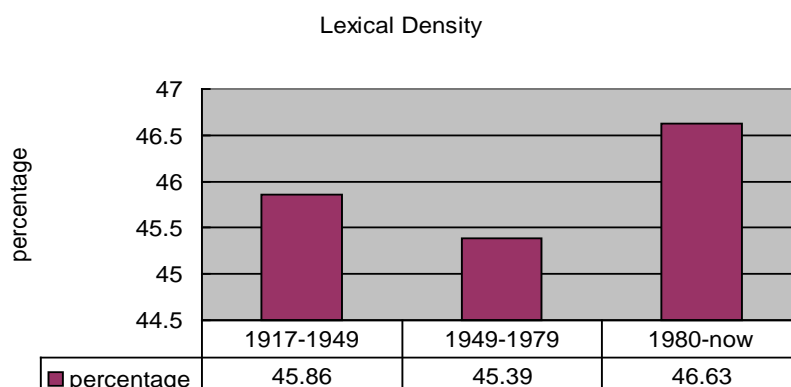|	1917-1949	1949-1979	1980-now
percentage	45.86	45.39	46.63

Figure 1. Figure of lexical density

The value of the lexical density of 1949-1979 is relatively low compared with the other two values, The operational translation norm of this period is simplification. Therefore, take it as the reference object, the percentage of the first period shows a feature of complication. Later in the third period, the percentage raises, and shows a relatively complication feature.

In order to get the reason for the changes, the background history information behind the translation of the novel needs to be looked at. In 1954, Shi Heng's version is under the very left ideology government, thus literature and translation, as the tool of politics and ideology, could not be free in itself either. The choice of words is very cautious, thus easy and repeated words are often used. After 1980s, the translators are free to express themselves, they have the freedom to choose the themes(that's why there emerges many translation versions of *The Scarlet Letter* which is originally viewed as a taboo since the theme of it is adultery) and the choice of multiple words. So the message contained in the texts increases and reflected in the percentage of lexical density, it increases.

Syntactic Features. By using the tool WordSmith 4.0, the average length and average sentence length can be calculated simultaneously, and the concrete figures are shown on the following table.

Table 2: Average Word Length & Average Sentence Length

	Average word length	1-letter words	2-letter words	3-letter words	4-letter words	Average Sentence Length
Zhang 1934	**1.22**	44,674	31,890	887	198	**44.56**
Fu 1937	1.44	47,471	32,308	1,158	280	74.51
Yang 1942	1.44	48,550	33,037	1,212	230	83.05
Shi 1954	1.48	41,995	33,478	986	369	79.88
Hu 1991	1.49	41,148	31,969	975	884	75.10
Zhou 1991	1.46	45,282	31,031	1,152	769	**235.70**
LiDong 1996	1.49	39,331	30,731	1,063	527	99.95
LiYanbo1996	1.48	39,654	29,237	1,113	608	74.14
Yao 1998	1.52	39,662	33,642	1,077	1,019	75.53
Feng 1999	1.53	38,367	33,672	1,198	804	70.30
Fang 1999	1.97	39,032	36,145	1,944	907	84.20
Xie 2000	1.47	43,570	30,871	944	792	73.42
Xiong 2000	1.50	38,570	29,791	1,017	886	**66.98**
Huang 2002	1.51	40,077	32,847	1,254	765	74.84
Dong 2002	1.49	40,646	30,719	1,343	726	77.93
Wang 2002	1.46	39,739	28,770	865	481	77.66
Zhou 2002	1.48	42,809	31,610	1,489	764	109.77
Bai 2003	1.47	43,708	31,792	1,205	840	75.75
Liu 2003	1.50	38,259	29,635	1,032	826	69.46
He 2004	1.50	39,339	31,959	957	775	76.05
Jia 2004	1.51	38.173	29,744	1,282	979	174.47
Wang 2006	1.48	39,724	30,461	910	816	69.22
Zhao 2007	1.47	35,246	24,706	882	649	87.13
Su 2007	1.54	39,259	35,307	1,683	921	99.24

From the above chart, we can conclude that the average word length of each of the translations is more or less the same around 1.5 except for Zhang Menglin's which is only 1.22. And because of the different choice of words of each translation, the average sentence length also varies. In 1934, the average sentence length of Zhang Menglin's version is only 44,56, the reason for the appearance of short sentences may somewhat due to the retained classical Chinese, and the influence is vernacular movement is just infiltrating gradually. So this translation is set aside for further studies, and therefore, the average sentence length of the first period is 78.78, in 1954, it is 79.88. After 1980s, the average sentence length is 85.79.

Average Sentence Length

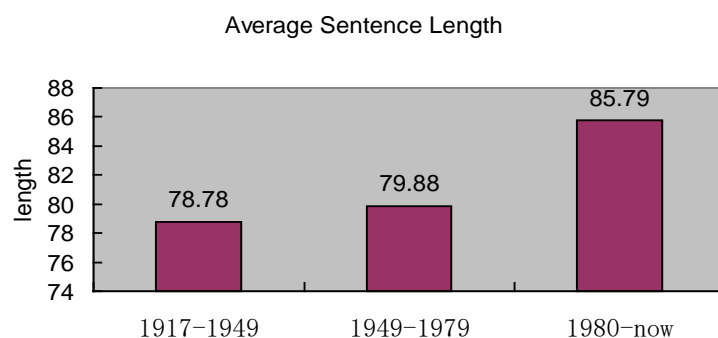| length	78.78	79.88	85.79
	1917-1949	1949-1979	1980-now

Figure 2. Average sentence length

On average, the translations tend to use longer sentences as the years goes by. And after 1980s, the sentence length reached the climax. And we call see from the table that the number of 3-letter and 4-letter words increased greatly. In a sense, the longer sentences manifest in syntactic level is the performance of explicitation feature, which means translators tend to produce more clarified target texts by adding descriptives, attributives, and elaborations in the target texts in order to make the versions more acceptable to the target audience.

Therefore, the operation norms of the other two periods shows the feature of implication relatively. Actually, the increase of sentence length is because of Europeanization. Before the beginning of European influence, long and complicated sentences with many embedded clauses or complex modifiers were quite rare in Chinese. Most traditional Chinese text was made up of a succession of short and syntactically simple phrases and sentences. After May 4th Movement, the Chinese language has experienced significant changes under the impact of the external influences like the big changes in social life and language contact through translation. A large number of Europeanized constructions and expressions have come into Chinese with the aid of translation, therefore forming a typical phenomenon in Modern Chinese - Europeanized Grammar. The Europeanization of Chinese grammar is one of the great events in history. And the evolution of the Vernacular Chinese after the May 4th Movement into today's Modern Chinese is virtually a process of constant Europeanization. Some early translators and writers, who hoped to improve the expressiveness of Chinese through introducing foreign language structures and expressions, produced deliberately a lot of lengthy and complicated Chinese sentences in their translations and writings. Their efforts have caused the appearance of a large number of lengthy sentences in Modern Chinese. And this kind of complicated sentence pattern is still playing an increasing influence on today's Chinese.

Conclusion

By building the monolingual corpus of the specific novel — *The Scarlet Letter*, the author uses corpus linguistic tools to investigate on the translations which are actually "The Third Code". The translations are classified into three historical periods according to the division of contemporary translation history, say, 1917-1949, 1949-1979, 1980s-now. The research has explored the translation norms of different translations in different historical periods, and some findings have been conducted. In the lexical level, by researching lexical density, the author gets that the translations of *The Scarlet Letter* is not dense texts, and the translation norm revealed in the three historical periods, say, 1917-1949, 1949-1979 and 1980-now are complication, simplification and complication respectively. So the diachronic changes of the norms are from complication to simplification and then complication again. In the syntactic level, by researching the average sentence length, the author gets the conclusion that the translators tend to use longer sentences as time goes by, say, from implication to explication, and this is because of the longer and multi-layer attributes, more use of idiom etc.

References

[1] Tang Jie, Shu Qizhi. Manipulation of Words and Different Images in the Eyes of Translators: from the Perspective of Social and Historical Background [J]. Journal of Beijing Electronic Science and Technology Institute, 2006, 14(3): 86-90.

[2] Xie Jiacheng. Self-built English & Chinese Parallel Corpus in Translation Study [J]. Journal of Yangtze University (Social Sciences), 2007, 30(1): 101-103.

[3] Cao Xiying. The Effect of Cultural Difference on Translation—Analysis of the Cultural Difference in Two Chinese Versions of The Scarlet Letter [J]. Journal of Xihua University (Philosophy & Social Sciences), 2006(1): 110-112.

[4] Cheng Shaohua. A Comparison between the Three Translation Versions of *The Scarlet Letter* in the Transplant of Symbolism and Figuration [J].2007(2): 47-50.