

# A Corpus-Based Comparative Study of Translation Universals in Two English Translations of *Li Sao*\*

CHANG Yan

School of Foreign Languages, Wuhan University of Technology, Wuhan, China, 430070

The present study chooses *Li Sao*, one of the representatives of traditional Chinese *Chu* culture, as the research subject. Through establishing two small monolingual corpora, a comparative analysis of Yang Xianyi's translation and David Hawkes' translation was made from the perspective of translation universals on three levels, namely lexical level, syntactic level and textual level. The results show that two translations both reflect the translation universals to some extent, but Hawkes' translation is more consistent with the translation universals in comparison with Yang Xianyi's on the whole.

*Key words:* *Li Sao*, corpus-based, comparative study, translation universals

## Introduction

As one of the representative works of *Chu* culture in Chinese spiritual civilization, Qu Yuan's *Li Sao* has made remarkable achievements in artistic creation. It is a realistic romantic lyric poem, especially in the shaping of the protagonist and the description of the characteristics of images with exaggeration of romanticism. The citations of a large number of myths and legends, together with the depiction of various flowers, birds and other images, broaden the imagination of the entire poem and strengthen the romantic charm of *Li Sao*. As a representative work, the style of *Li Sao* is also named the *Sao* style by later generations, exerting a far-reaching influence on the development of literary genre. Many scholars have studied the translation of *Li Sao* from different perspectives, such as the perspective of intertextuality, the theory of iconicity, the comparison of literary features, and the hermeneutic perspective and so on. In contrast, there are few articles making corpus-based translation study of *Li Sao* with quantitative method. Thus, through the construction of two monolingual corpora, a comparative analysis of Yang's and Hawkes' translation was made to provide reference for the translation of other traditional Chinese classic poems.

## Research Methodology

### Corpus Collection, Construction and Tagging

Firstly, the two translation versions are scanned and organized into txt files, and after manual verification, two small monolingual corpora are established, namely Corpus of Yang Xianyi's Translation (CYT as its

---

\* **Acknowledgements:** This paper is supported by the Excellent Dissertation Cultivation Funds of Wuhan University of Technology (2018-YS-090).

CHANG Yan, Master of Art, Master Student, School of Foreign Languages, Wuhan University of Technology.

abbreviation) and Corpus of David Hawkes' Translation (CHT as its abbreviation). And the British National Corpus (BNC), as the comparable corpora, is a 100-million-word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the late twentieth century, which has the authority and comprehensiveness as the representative. After the construction of the corpus, the part-of-speech tagging is performed through Free CLAWS web tagger on the two small corpora, in order to conduct in-depth multidimensional comparative research from the perspectives of lexical, syntactic and textual level.

### **Corpus Analysis Software**

Corpus data-driven belongs to quantitative analysis, which derives conclusions from specific data and has repeatable verification. So it is objective and effective, which greatly overcomes the subjectivity and arbitrariness of translation studies, and it is a powerful supplement to qualitative analysis. According to different research parameters at lexical level, syntactic level and lexical level, this study obtains quantitative data with the help of corpus tools such as Antconc 3.2.2.1, WordSmith 6.0, Wordless 1.3.0, and then carries out qualitative comparative analysis.

## **Data Results and Analysis**

### **On the Lexical Level**

Vocabulary selection is a process in which translators continuously make decisions during the translation process. Not only can they largely reflect the different translators' styles and methods adopted, but also through the development and application of corpus translation methods, lexical-level study can reveal the translation universals and tendencies of translation language such as explicitation as well as simplification. Therefore, the study at the lexical level is essential for translation studies. The present study makes a comparative research on the simplification degree of the CYT and CHT mainly through the perspectives of type/token ratio and word length, lexical density.

Baker (1996) states that simplification means that the translator unconsciously simplifies language, or information, or both. The earliest linguist to verify simplification of translation language through corpus was Leviosa, whose research confirms that the vocabulary range of translated text is narrower than that of untranslated text, and the vocabulary density of translated text is lower, also the average sentence length of translated text is shorter than the average sentence length of non-translated text (2002).

#### **Type/token ratio and word length**

Baker (2000) points out that type/token ratio (TTR) is directly proportional to the lexical diversity of texts. The higher of this ratio, the translation language has the greater lexical variety and richer in words. On the contrary, it indicates that the vocabulary used by the translator is lesser and the lexical variety is small. Due to TTR is easily affected by the length of the text or the corpus capacity, Scott (2006) and other researchers propose a formula for calculating vocabulary richness-the standard type/token ratio (STTR), which uses 1000 words as the unit to calculate the type/token ratio per 1,000 words, and finally calculates the average value that has the referential value.

Table 1

*The Type/Token Ratio of CYT and CHT*

	CYT	CHT	BNC
<b>Type</b>	1101	1208	166,962
<b>Token</b>	2916	3510	102,467,488
<b>TTR%</b>	37.7	34.4	0.16
<b>STTR%</b>	50	48	41.2

From Table 1, it is clearly to see that the TTR of Yang's translation is higher than Hawkes' translation, which indicates the Yang's translation language has higher lexical variety than Hawkes' although the overall length of Hawkes' translation is longer. But both the STTR of CYT and CHT are higher than BNC, which may relate to the special genre and full of various images of the source text. Compared with Yang's translation, the wording of Hawkes' translation is relatively simpler with few lexical varieties, which is relatively close to that of native speakers.

Table 2

*The Word Length of CYT and CHT*

	CYT	CHT	BNC
<b>Word length (mean)</b>	4.38	4.18	4.66

Table 2 indicates that the average word length of Yang's translation is a little longer than Hawkes', which shows the reading difficulty of Yang's version is slightly higher than that of Hawkes'. But both of them are a little bit lower than the word length of BNC, which indicates that the words used in both CYT and CHT are in normal level as English native speakers.

**Lexical density**

In English, lexical words refer to words with stable lexical meaning, including nouns, adjectives, verbs and most adverbs; while functional words refer to words that do not have stable lexical meaning or mainly have grammatical functions due to their vague meanings, mainly including pronouns, prepositions, conjunctions, articles and auxiliary verbs (Hu, 2007). This study adopts the formula of lexical density proposed by Stubbs (1986): lexical density = the number of lexical words/the total number of words\*100%. The more lexical words in the text, the greater the lexical density. Hu (2007) points out that vocabulary density can reflect the carrying capacity of text information. The higher vocabulary density, the higher frequency of using lexical words, the greater amount of text information, and the higher difficulty of reading; otherwise, the opposite situation will occur. Through writing regular expression in Antconc in order to process the CYT as well as CHT with POS tagger, Table 3 clearly shows the lexical word and its proportion in both CYT and CHT.

Table 3

*The Total Lexical Word and Proportion of CYT and CHT*

Lexical Word	CYT	Proportion%	CHT	Proportion%
<b>N</b>	689	23.6	711	20.3
<b>V</b>	488	16.7	500	14.2
<b>AJ</b>	252	8.6	236	6.7
<b>AV</b>	161	5.5	254	7.2

<b>Total</b>	1590	54.4	1701	48.4
--------------	------	------	------	------

It is clear to see that though the total number of lexical words in Yang's translation is lesser than Hawkes', the proportion of lexical word in CYT is higher than CHT due to the total tokens in Hawkes' translation is more than Yang's translation. The proportion indicates that CYT has a higher lexical density and it carries more text information than CHT, and its degree of reading difficulty is higher than CHT.

### On the Syntactic Level

Blum-Kulka (1996) believes that a successful translation requires multiple approaches to processing the discourse, thus the translation version may be longer than the source text, but the cohesion will be clearer. From Baker's view, the theory of explicitation concerns the tendency in translations to spell things out rather than leave them implicit (Baker, 1996) Therefore, the explicitation on syntactic level in this study mainly focuses on the conjunctions and personal pronouns in CYT and CHT.

With the help of Antconc, Table 4 shows the conjunctions used in CYT and CHT.

Table 4

#### *The Conjunctions Used in CYT and CHT*

<b>Conjunctions(CJ)</b>	<b>CYT</b>	<b>Proportion%</b>	<b>CHT</b>	<b>Proportion%</b>
<b>Coordinating Conjunctions(CJC)</b>	112	3.8	176	5
<b>Subordinating Conjunctions(CJS)</b>	47	1.6	56	1.6
<b>Conjunction That (CJT)</b>	13	0.4	23	0.6
<b>Total</b>	172	5.8	255	7.2

On account of the special genre of *Li Sao*, few conjunctions are used in the source text. The source text has the unique characteristics of Chinese ancient poetry, namely, compactness and conciseness. Because the Chinese language attach more importance on parataxis, while English focuses on hypotaxis. Thus, in the translated language, both Yang Xianyi and Hawkes use different conjunctions to express the inner relations of the sentences implied in the original text, so that the target readers can understand the logical relations clearly. According to the proportion showed in above table, it can be seen that Hawkes uses more conjunctions than Yang Xianyi to make the cohesion within sentence clearer. Thus, the degree of explicitation of Hawkes' translation is higher than Yang's and compared with Yang's translation, Hawkes's translation is easier for target readers to understand the content of the poem.

The frequency of personal pronouns used in CYT and CHT are showed in Table 5 through Antconc. It can be obviously seen that the most frequent used personal pronoun both in Yang's translation and Hawkes' translation is *I*, which is because the source text is a self-reporting lyric poem. Since the subject is not a necessary element in Chinese grammar, which cannot be omitted in English expressions, so few personal pronouns appear in the original poem. However, for the sake of grammar and logic, both Yang Xianyi and Hawkes add many personal pronouns to complete the implied subject and subordinate relation of the sentences. At the same time, compared with Yang's translation, more personal pronouns are used in Hawkes' translation, which indicates that its explicitation degree on syntactic level is higher than Yang's translation.

Table 5

*The Frequency of Personal Pronouns CYT and CHT*

Personal Pronouns (PNP)	CYT	Proportion%	CHT	Proportion%
I	103	3.5	122	3.5
me	11	0.4	32	0.9
mine	1	0.03	0	0
they	26	0.9	13	0.4
them	3	0.1	8	0.2
thou	6	0.2	0	0
you			15	0.4
thee	3	0.1	0	0
he	6	0.2	19	0.5
him	4	0.13	9	0.3
she	4	0.13	7	0.2
her	6	0.2	6	0.2
we	4	0.13	4	0.1
it	5	0.17	15	0.4
<b>Total</b>	<b>182</b>	<b>6.2</b>	<b>250</b>	<b>7.1</b>

**On the Textual Level**

Normalization means that “the translator consciously or unconsciously eliminates the uniqueness of the source language during translation, making the translation more in line with the standard of the target language” (Hu, 2004). Through research, Vanderawera (1985) finds a tendency to revise punctuation, lexical choice, style, sentence structure and textual organization. All these manipulations tend to textual conventionality, apparently approved by target audience. In order to manifest normalization on the textual level, this study mainly discusses the translation of proper noun, culture-loaded words as well as the punctuation and some relative words.

As Table 3 shows, both of the proportion of nouns in Yang’s translation and Hawkes’ translation is the highest one among the lexical words. This may be due to a large number of cultural loaded words and proper nouns such as flowers and birds, myths, names and places in the source text. Part of the proper nouns are showed in the following table.

Table 6

*Some of Proper Nouns Used in CYT and CHT*

Proper Nouns	CYT	CHT
帝高阳	Ancestry renowned	lord Gao Yang
启	a Monarch’s Son	Qi of Xia
夏桀	the Prince	Jie of Xia
九天	Celestial Spheres	the ninefold heaven
《九辩》	Odes divine	Nine Changes
《九歌》	Odes divine	Nine Songs

In the source text, Qu Yuan uses a large number of examples involving the name of the ancients, some nouns with special meaning in Chinese ancient culture. For example, 帝高阳 refers to one of the five emperors in ancient times and he was one of the tribal leaders making outstanding contributions to human beings in the later period of primitive society. Thanks to different cultural backgrounds, Yang translates it into *Ancestry renowned*

to show Qu Yuan's nobility of birth instead of adopting transliteration. In the same way, Yang conceals all the names involved in myths or allusions and abandons their literal meanings in the whole translation to eliminate the obstacles of readers' understanding of a foreign culture. In contrast, when Hawkes dealing with this kind of proper nouns, he tries to keep the alienation of a foreign culture and uses transliteration as well as annotation behind the translation in order to convey the Chinese culture implied. Therefore, at the aspect of normalization on the textual level, Yang's translation is easier for target readers to understand the content, but Hawkes' translation is help for introducing the Chinese culture implied behind the poem.

Since the genre of the source text belongs to a kind of poetry, named as *Sao* Style creating by Qu Yuan, its sentence structure is neat and its punctuation is standard. Table 7 shows the total punctuation used in the source text is 381, which is much lesser than in these two translations. The reason may attribute to the compactness and conciseness of the poem with a lot of information contained. However, it is difficult for translating the source text in accord with the exact number of sentences in English because punctuation and relative words are needed in English for conveying as much information as the source text meanwhile keeping the translation grammatically correct. Thus, both Yang and Hawkes add some relative words and punctuation in their translation. Therefore, in terms of textual level, both CYT and CHT have the characteristic of normalization, but the degree in the Hawkes' version is higher than that in Yang's version.

Table 7

*The Punctuation and Main Relative Words Used in CYT, CHT and ST*

Punctuation	CYT	CHT	Source Text (ST)
,	213	179	186
。	119	147	157
?	25	26	21
;	77	44	0
:	5	31	8
!	2	9	9
when	3	8	
which	1	2	
that	17	34	
who	10	5	
where	10	6	
<b>Total</b>	<b>482</b>	<b>491</b>	<b>381</b>

There are two kinds of situation leading to more sentences and connective words added in the translated language, one is breaking down an informative sentence in source text into two sentences with punctuation, the other is using a complex sentence with a connective word.

### Conclusion

Based on the corpus linguistics and statistical analysis research methods, the present study compares the Yang Xianyi's and David Hawkes' translation of *Li Sao* with a multidimensional research from lexical level, syntactic level, and textual level. Combined with quantitative as well as qualitative analysis, this study comparatively analyzes and carries out a descriptive analysis about the linguistic features and translation universals of these two translations. Here are the conclusions the study reached: Firstly, simplification is reflected

in both Yang Xianyi's and Hawkes' translation on the lexical level. But compared with Yang's translation, the lexical variety of Hawkes' translation is lower, its word length is shorter and its lexical density is lower, thus the degree of simplification of Hawkes' translation is higher than Yang's. But the STTR of these two translations is higher than BNC. Secondly, on the syntactic level, both Yang Xianyi and Hawkes add conjunctions and personal pronouns to make the implied meaning clearer and grammatically correct, which shows the explicitation of the translated language. By adding more conjunctions and personal pronouns, the degree of explicitation of Hawkes' translation is also higher than Yang Xianyi's. Thirdly, on the textual level, Hawkes adopts the foreignization strategy in the translation of proper nouns and culture-loaded words, which deviated from normalization in translated language. While Yang Xianyi adopts the domestication strategy and tries to eliminate the reading barriers. However, when dealing with the informative sentences in the source text, they both choose to use simple sentences or clauses. On the whole, both Yang Xianyi' translation and David Hawkes' translation reflect the translation universals to some extent, but Hawkes' translation is more consistent with the translation universals in comparison with Yang Xianyi's.

### References

- Baker, M. (1996). Corpus-based translation studies: The challenges that lie in ahead. In Harold Somers (Ed), *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*. Amsterdam: John Benjamins.
- Baker, M. (2000). Towards a methodology for investigating the style of a literary translator. *Target*, (2), 241-266.
- Blum-Kulka, S. (1996). Shifts of cohesion and coherence in translation. In J. House & S. Blum -Kulka (Eds.), *Interlingual and intercultural communication: Discourse and cognition in translation and second language acquisition studies*. Tübingen: Narr.
- Hawkes D. (1985). *The songs of the south—An anthology of ancient Chinese poems by Qu Yuan and other poets*. Harmondsworth: Penguin Books Ltd.
- Hu, X. Y. (2004). Corpus translation studies and translation universals. *Shanghai Journal of Translators for Science and Technology*, (4), 47-49.
- Hu, X. Y. (2007). A corpus based study on the lexical features of Chinese translated fiction. *Foreign Language Teaching and Research*, (3), 214-220.
- Laviosa, S. (2002). *Corpus-based translation studies*. Amsterdam-New York: Rodopi.
- Scott, M., & Tribble, C. (2006). *Textual patterns: Keyword and corpus analysis in language education*. Amsterdam & Philadelphia: John Benjamins Publishing Co.
- Stubbs, M. (1986). Lexical density: A computational technique and some findings. In M. Coulter (Ed.), *Talking about text. Studies presented to David Brazil on his retirement*. Birmingham: English Language Research, University of Birmingham.
- Vanderawera, R. (1985). *Dutch novels translated into English: The transformation of a "minority" literature*. Amsterdam: Rodopi.
- Yang, X. Y., & Dai, N. (2016). *Selected elegies of the state of Chu* (pp. 2-39). Beijing: Foreign Languages Press.