

Corpus-based Analysis of Linguistic Features of English Versions of Watson's Classics

Lili Sun

College English Teaching Department
Zaozhuang University
Zaozhuang, China

Abstract—By use of the small self-built corpus and the use of corpus research methods, this study analyzes the linguistic features of the translation of Watson's classics from three aspects: vocabulary, syntax and textual cohesion. Through the specific language phenomena, it is concluded that the language of the translation of Watson's classics is concise and easy to understand. The language used is graceful and fluent, so it is more readable.

Keywords—vocabulary; syntax; discourse; corpus

I. INTRODUCTION

In recent years, corpus-based studies on translation have become more and more popular. With the help of corpus linguistics, many scholars, guided by literature, linguistics and translation theories, have made a comprehensive and systematic observation. They have generalized the corpus on the basis of a large number of real data and with the help of corpus linguistics research methods. The results of the study are more authentic and scientific, making translation research more specific and objective.

Watson, an American sinologist and translator, has translated more than 20 works into Chinese classics, and has formed his own style of language translation. His translation is simple and elegant. The translation language is graceful and fluent, and the readability is very strong. His language is simple, plain and vivid. He has its own unique characteristics in language form and artistic style. This study, with the help of a corpus, discusses the linguistic features of Watson's English translation from three aspects: vocabulary, syntax and textual cohesion. A small corpus was built for this purpose. The corpus mainly includes the excerpts of Mozi Volume 1, Section 8, Section 9; Section 9 of Xunzi; Section 5 and Section 8 of Hanfeizi.

II. LEXICAL LEVEL

Vocabulary is the center of language. Michael F. Graves summarized the main contents of English vocabulary research over the past 100 years. The study of vocabulary can be summarized as "vocabulary knowledge is one of the best indicators of language competence; vocabulary knowledge indicates language reading comprehension ability; lexical difficulty limits the readability of the text; vocabulary is the main focus of English language research."

In this study, the word length, word density, word frequency and classifier are used to analyze the lexical features of Watson's English translation. With the help of corpus-related software, accurate statistics of figures and sentences are obtained to interpret the linguistic features of Watson's English translation.

A. Average Word Length

One hundred and fifty years ago, Lord proposed in his book that word length could be used as a measure of the author's personal writing style. Word length = word number (excluding punctuation) / total number of words. Word length refers to the total number of letters contained in a word. It helps us to determine whether the common words are formal or informal, simple or compound. Previous studies have shown that the average word length of a text is about 4 letters, if less than 4 letters, that the article belongs to the spoken language; if more than 4 letters, that the language of the article is more complex, more difficult words. In this paper, the average word length in the self-built corpus is calculated by WordSmith 4.0 software with reference to the average word length in the three corpora of spoken English, fiction and academic English. (See "Table I")

TABLE I. MEAN WORD LENGTH

styles	Self-built	conversation	novel	academic
Mean word length	4.59	3.78	4.27	4.91

The above corpora are all from the corpora of Beijing Foreign Studies University. As the table shows, academic English has the longest word length, followed by News English, and the shortest is spoken English corpus. Generally speaking, the more complex the words are, the more obscure and difficult the semantic expressions are. In everyday communication, we tend to use short words to communicate. The average word length of Watson's English translation is 4.59, which is close to that of fiction and slightly longer than that of spoken English. The data show that the long and difficult words used in the Watson's translation are fewer and the length of the words is shorter. This phenomenon shows that the words in Word son's translation are less difficult. The language is simpler and easier to understand, and the reading of the article is less difficult.

According to the specific statistics on the length of words in Watson's translation, the number of words with three letters is 5,015, accounting for 20.7% of the total words. The number of words with two letters is 4,208, accounting for 17.4% of the total words. The number of words with four letters is 3,633, accounting for 15% of the total words. It can further show that most of the words in Watson's translation are simple words, so it is less difficult to understand.

B. Vocabulary Difficulty

The lexical difficulty of a text refers to the distribution of words in each level according to the reference vocabulary, with the help of Antword profiler software and its own reference vocabulary. The software Antword profiler comes with a vocabulary of three levels. The first-level vocabulary consists of 1000 words with the highest frequency in English. The second-level vocabulary consists of the 1001-2000 most frequent words in English. The first two glossary came from the first 2000 words of the general words of English vocabulary formulated by West (1953). The third vocabulary corresponds to Coxhead's (2000) Academic vocabulary, which includes 570 commonly used academic vocabularies in Universities in addition to the 2000 most commonly used ones. The results of word searching by software are shown in "Table II" below.

TABLE II. LEXICAL DIFFICULTY

	LEVEL LIST	TOKEN	TOKEN%	TYPE
1	1899	78.18	1404	37.3
2	1212	5.01	512	13.6
3	935	3.87	325	8.63
0	3128	12.94	1523	40.46
Total	24174		3764	

The chart above shows that there are 24,174 characters in the self-built corpus, of which 18,899 words belong to Level 1, accounting for 78.2% of the total number of words in the text; 1,404 words belong to Level 2, accounting for 37.3% of the text and 740 words belong to the text, accounting for 26.3% of the total word family. The distribution of Level 2 is analogous. Level 3 has a small total of 935 words, accounting for 3.8% of the total. This shows that Watson's language focuses on simple communicative language and has many forms of dialogue, rarely involving academic expression. Level 0 is a word that does not appear in Level 1, Level 2 and Level 3. Through the analysis of the corpus, it is found that the names of persons, places, dynasties and units of measurement are frequently used in Watson's translation versions. Most of these words are literally translated into Chinese Pinyin, and annotated. These words are not included in the general English or academic English, so they are not found in Level 1-3. For example, Mount Tai; North Sea; Qin and Chu; Duke Wen of Teng; Yao; Fu Lake; Yao, Shun, Yü; T'ang and King Wen; the Hsia, Shang, and Chou Dynasties. From the above data, the vocabulary of the Chinese version corpus is mostly concentrated in the category of 1000 words, which belongs to the most commonly used English words at the level of symbols, morphology and family. Therefore, the overall reading difficulty of the Chinese version is not high.

C. High Frequency Words

With the help of computer, we can easily generate word frequency table. The word on top of the table is high-frequency words. As shown in the following figure, by running AntConc 3.2.0 software, we can get the words with high frequency in the self-built corpus. The words with the highest frequency are all the functional words, such as the, of, to, in, a and so on. (See "Table III")

TABLE III. HIGH FREQUENCY WORDS

Rank	Freq.	Word	Rank	Freq.	Word
1	1404	the	16	166	People
2	1052	and	17	151	that
3	816	of	18	143	them
4	729	to	19	143	who
5	392	in	20	133	the
6	360	is	21	129	it
7	338	a	22	125	for
8	304	be	23	114	do
9	286	will	24	112	ruler
10	248	he	25	112	with
11	245	his	26	109	men
12	219	their	27	106	as
13	210	not	28	100	no
14	209	are	29	100	this
15	189	they	30	99	have

From observation of the vocabulary in the middle, the function words are rarely seen and basically are lexical words. The frequency of the third person pronouns and the subject pronouns he, his, his, they is high. First, the frequency of two-person pronouns I, you is almost zero. This phenomenon is mainly due to the high frequency of the first and two-person pronouns in spoken English. However, the books of Mozi involve more Mohist military thoughts, political thoughts and educational thoughts. These thoughts are stated by putting facts and reasoning. The mode of expression, rarely used to express personal views of I, you and other words, because of the political and military fields, such as king, ruler, government, state, ministers, officials and other related words appear more frequently.

Based on the previous explicit studies of the subject of personal pronouns in the translated literary style (Huang Libo 2008; Ren Xiaohua 2015; Wang Xinxin 2015; Zhao Shikai 2011; Zhao Hong2002), the study of Chinese literary texts and their corresponding English translations reveals that the use of third-person pronouns is prominent after the Chinese text is translated into English text, and the English version is also prominent. The use of the third personal pronouns is at least two to three times higher than that of the original Chinese. Chinese pays attention to parataxis in the expression of semantics, relying on the inherent meaning of sentences and words to express logical relations, while English pays attention to hypotaxis, with the help of cohesive words to make the text coherent. Explicitness of personal pronouns means that the translator uses cohesive devices to make the translation more readable. Flexible translation of personal reference is also an expression of the translator's accurate and standard translation. The use of personal pronouns can make the translation accurate, clear in meaning

and avoid ambiguity. It is a reflection of the translator's translation ability. In the process of translation, the translator can make the reader understand the original text by changing the way of language expression or increasing the corresponding background knowledge to understand the deep social and cultural meanings conveyed in the literary works. The data show that Wordson can vividly and concretely express the relationship between the characters in the classics and illustrate the translator. The translation ability is very excellent and its translation work is highly readable.

III. SENTENCE LEVEL

A. Sentence Length

Sentence length refers to the total number of words in a complete sentence. Bulter (1985) classifies sentences into three categories according to their length: short sentences (1-9 words), medium-length sentences (10-25 words) and long sentences (more than 25 words). With the help of WordSmith 4.0, we can easily get the sentence length in the self-built corpus. According to the figures, the average sentence length of the Chinese version is 22.6 words, belonging to the medium-length sentence, which is close to the sentence length of the English native speaker BROWN corpus (23.37), and there are many parallel sentences connected with 'and' and 'but'. This shows that Watson always insists on translating Chinese classics in popular and natural modern English. The language of the translated version is not simply to reflect the ancient features of the original text instead, to present the contents of classics to the readers in modern English. The translation is highly readable.

B. Sentence Types

"Sentences include simple sentences and complex sentences. Simple sentences consist of a separate sentence, and complex sentences consist of one or more sentences as part of it. Complex sentences include coordinate sentences and compound sentences." (Quirk etc. 1972:719) Using their viewpoints, this study randomly makes the following statistics on the frequency of simple sentences, compound sentences and complex sentences in 100 sentences in the self-built corpus:

TABLE IV. SENTENCE TYPES

types	total	rate
Complicated sentences	9	9%
Simple sentences	18	18%
Complex sentences	73	73%

"Table IV" shows that in the English translation of classics, compound sentences appear most frequently, followed by simple sentences and less complex sentences. Watson used fewer complex sentences in English translation, which made the sentences not lengthy and be concise. There are fewer simple sentences in the self-built corpus, because too short sentences cannot fully convey semantics, therefore there are fewer simple sentences. Compound sentences are most frequently used. In order to convey the original ideas

accurately and appropriately, translators use attributive clauses (who), time adverbial clauses (when) and conditional adverbial clauses (if). The use of clauses can make the meaning of sentences more clear and the sentences more fluent. Take the following translation as an example:

- Example (1): Let us suppose there is a rich man who has built a high wall all around his house.
- Example (2): When the wall is finished and plastered with mud, he pierces it with only one gate.
- Example (3): Then, if a thief steals in, he may shut the gate by which the thief entered and set about searching for him, confident that the thief has no means of escape.

The above example sentences are a small paragraph of the translated text randomly selected by the author. Through analysis, the translator uses a variety of clauses guided by the conjunctions 'who, when, if, by which and that'. The use of the clauses enables the sentences to express the past and the future, and to convey the ideas of the original text in a logical way, so that the readers can easily understand what principle Mozi should follow in explaining the wise monarch's recruitment of wise scholars.

C. Rhetoric

Rhetoric is the use of art and language, which has a convincing effect. In order to persuade the other party to accept his ideas, Mozi used a lot of rhetoric to express his military, political and educational concepts. The most frequently used rhetorical devices in translation are simile, parallelism, rhetorical questions and emphasis.

1) *Simile*: Simile is a comparison of things with common characteristics. The commonly used symbols are 'like, as, seem, if, as though, similar to', etc. Through the analysis of the translation, simile with 'like' as the symbol is used more frequently.

- Example (4): Because the rich man, like the ruler, has control of the vital point.
- Example (5): Never faltering, never failing, Brilliant like the sun, Bright like the moon, Constant as heavy
- Example (6): Never failing, Brilliant like the sun, Bright like the moon, Constant as heaven and earth.
- Example (7): Gradual as he slimming moon, your enhancing like a slow-spreading heat.
- Example (8): Such a desire by acting in the way you do is like climbing a tree in search of a fish."

In the above examples, 'rich man' is compared to 'ruler'; 'brilliant' is compared to 'the sun'; 'bright' is compared to 'the moon'; 'such a desire' is compared to 'climbing a tree in search of a fish'. All these metaphors make statements and opinions more understandable.

2) *Parallelism*: Parallelism refers to the arrangement of sentences or sentence elements with the same or similar structure, the same mood and the closely related meaning,

so as to enhance the sentence pattern. In the example sentence, ‘the unrighteous shall not be...’; ‘there were none who ...’; ‘those who... Repeated occurrence makes the sentence pattern the same, make the semantic expression more concise, consistent with the original style.

- Example (9): The unrighteous shall not be enriched, the unrighteous shall not be exalted, the unrighteous shall be no kin to us, the unrighteous shall not be our intimates!
- Example (10): ...there were none who were not unfailingly cautious and respectful, none who did not encourage and strive with each other in honoring virtue.
- Example (11): Those who were capable of ordering the state were employed to order the state; those who were capable of heading a government bureau were employed as heads of bureaus; and those who were capable of governing an outlying district were employed to govern the outlying districts.

3) *Antithesis*: Antithesis is a rhetorical way of comparing sentences with opposite meanings. ‘Stay’ and ‘venture forth’ in Example 1, ‘worth’ and ‘wicked’ in Example 2 and ‘rewarded’ and ‘punished’ in Example 2 are contrasted.

- Example (12): If they stay within their realm, their position will be secure, and if they venture forth to punish an enemy, they will be victorious.
- Example (13): If the worthy are not rewarded and the wicked are not punished, then there will be no way to encourage the worthy or put a stop to evil.

4) *Emphasis*: Emphasis sentences highlight the key points of the sentences, with distinctive primary and secondary, so that readers can maximize access to the main information. In translation, “it is...” plus emphasized part of the sentence structure is used more frequently.

- Example (14): It is gentlemen of true worth, therefore, who must act to assist and carry on the government.

5) *Quotation*: Quotation refers to the rhetoric method of using ready-made words in speech or writing, such as poems, aphorisms, idioms, etc., to express one's thoughts and feelings. The following example sentences directly quote the original words in the Book of Songs to provide strong arguments and persuasion for their own views and opinions.

- Example (15): The Book of Odes says: I admonish you to take thought for the needy; I teach you how to assign the titles; for who can take hold of something hot without first moistening his hand?

6) *Metonymy*: Metonymy refers to the rhetorical way of choosing the most prominent and obvious things to replace the name of a specific thing. A great deal of metonymy is used in the translation.

- Example (16): Therefore the sage kings of ancient times took great pains to honor the worthy and

employ the capable, showing no special consideration for their own kin, no partiality for the eminent and rich, no favoritism for the good-looking and attractive.

In the examples, adjectives such as “worthy”, “capable”, “eminent and rich”, “good-looking and attractive” are used instead of groups with such characteristics. Metonymy is used to make the language concise and vivid, to avoid repetition and redundancy, and to give people a fresh feeling.

7) *Rhetorical question*: Rhetorical questions express the meaning of negation or affirmation in the form of affirmative or negative interrogative sentences.

- Example (17): But why do they fail to perceive that honoring the worthy is the foundation of government?

Rhetorical questions include the definite meaning to be expressed in the questions, emphasizing the meaning to be expressed, making the original thought more clearly. Rhetorical questions are more powerful and emotional than general statements.

IV. DISCOURSE LEVEL

If the text itself is coherent, the text is naturally cohesive. They complement each other and are indispensable (Qin Xiubai, 2002). Chinese pays attention to parataxis in semantic expression, expressing logical relations by means of the internal meaning of sentences, while English pays attention to hypotaxis, expressing the internal relations of discourse by means of a large number of cohesive words, which makes the semantic expression clearer and the cohesion of discourse more natural. According to the statistics of AntConc software, the frequency of conjunctions in Chinese translation is higher, especially the frequency of ‘and’, ‘but’ and ‘when’, followed by the emergence of ‘if’, ‘but’ and ‘when’.

By searching the keyword, the context in which ‘and’ are used is as follows:

- 1 Master Mo Tzu 1 said: These days the rulers and high officials who govern the nation all desire
- 2 states to be rich, their population numerous, and their administration well ordered. And yet wha
- 3 umerous, and their administration well ordered. And yet what they achieve is not wealth but pove
- 4 actual fact, they fail to get what they seek and instead achieve what they abhor. Why is this?
- 5 this? Mo Tzu said: It is because the rulers and high officials who govern the nation fail to
- 6 administration will be characterized by weight and substance; but if it is poor in such men,
- 7 wishes to increase the number of skilled archers and chariot drivers in the state. One must set

In sentences 1, 5, 6, 7, ‘and’ is used to connect nouns that express juxtaposition, such as ‘rulers and high officials’; ‘weight and substance’; ‘archers and drives’ etc. In sentence

3, the word ‘and’ plays the role of connecting link, because the excerpted "Mozi" involves Mozi's political and military thoughts. In the process of explanation, there are many statements connecting the preceding and the following, so the frequency of use of the word ‘and’ is higher.

By searching the keyword, the context in which the word ‘if’ are used as follows:

- Employ the capable in their administration. *If* a government is rich in worthy men,
- He pierces it with only one gate. Then, *if* a thief steals in, he may shut the gate by
- a farmer or an artisan from the shops, *if* he had ability they promoted him. Such men
- would have no confidence in them; and *if* their orders were not carried out, the people
- uses act to assist and carry on the government. *If* the ruler can obtain the services of such gentle
- will be brought to a successful conclusion. *If* they stay within their realm, their position will
- their realm, their position will be secure, and *if* they venture forth to punish an enemy, they will
- become the leaders of the other lords. But *if* one knows only the policy to be adopted,
- What are these three principles? They are that *if* the titles and positions of worthy men are not
- enough, then the people will not respect such men; *if* their stipends are not generous, then the people

The word ‘If’ has appeared 180 times in the translated version. Through analysis, the word ‘if’ is used basically to express conditions and assumptions. The subject of the ‘if’ clause is ‘government’, ‘the thief’, ‘rulers’, ‘official’, ‘the worth’, ‘their orders’, ‘the title and positions’. When Mozi propagates Mohist military and political thoughts, he prefers to explain and disseminate them through specific image examples and assumptions. In the process of translation, Wordson maintained the style of the original text. After Mozi is translated into English, conjunctions such as ‘if’, ‘but’, ‘and’ were frequently used in the English version.

V. CONCLUSION

With the help of corpus linguistics tools and self-built small corpus, this paper studies the linguistic features of Wordson's translation version of Mozi from three aspects: vocabulary, syntax and discourse. Through the study of the use of word length, word frequency, class characters, vocabulary density, sentence length, rhetorical devices and conjunctions, it is found that the language of Wordson's translations is simple and easy to understand, and the linguistic form is closely linked. The style of the translated text is similar to that of the original text. It is highly readable to translate classical Chinese into modern English.

REFERENCES

- [1] Huang Libo. Appearance of Subjects of Personal Pronouns in English-Chinese Translation: A Corpus-based Study, *Foreign Language Teaching and Research*, 2008 (6): 454-459.
- [2] Ren Xiaohua. Explicitness of Personal Pronouns in Chinese Translation: A Corpus-based Study, *Journal of PLA College of Foreign Languages*, 2015(2):125-133.
- [3] Wang Xinxin. A Corpus-based Study on the Explicitness of Personal Pronouns in Literary Works: A Case Study of the English Version of *Fortress Besieged*. *Journal of ShangLuo university*, 2015 (3): 52-57.
- [4] Zhao Hong & Shao Zhihong. A contrastive study of anaphoric functions of third-person pronouns in English and Chinese, *Foreign Language Teaching and Research (Foreign Language Bimonthly)*, 2002(3): 174-179.
- [5] Zheng Jiewen & Zhangqian. *Mozi*, Nanjing University Press, 2010.
- [6] Feng Zhengbin & Lin Jiabin. Translation Strategies and Enlightenment of Wordson's Chinese Poetry into English, *foreign language teaching*, 2015(9):101-104.
- [7] Quirk, R., Greenbaum, S., Leech G. and Svartvik, J. *A Grammar of Contemporary English*. London: Longman: 729, 1972.
- [8] Fowler, R. *Style and Structure in Literature: essays in the new stylistics*. Oxford: Blackwell, 1975.