A Corpus-Based Comparative Analysis of Cohesive Devices in Two English Translations of *The Analects of Confucius*

Yuqiu Hou and Yu Sun

Abstract—The translation of discourse, as an important part of the translation field, is the application of the theory of discourse linguistics in translation studies. In text translation, cohesion theory has to be mentioned as an important branch and component of text linguistics, which has already received extensive attention at home and abroad and also shows that there are certain differences between the two languages in English and Chinese, since its first introduction of Halliday and Hasan in 1976. Because of this difference, the use of cohesive devices in translation also affects the quality of the versions of translation.

Moreover, in order to spread Chinese traditional culture and construct Chinese national image, the translation of Chinese classic works has gained a lot of attention of the whole society. Among these Chinese classic works, *The Analects of Confucius* is one of the most important and well-known works which has a great influence on Chinese thoughts and cultures. So in this paper, the author chooses two translation versions translated by Legge and Ku Hungming and uses the corpus statistics tool to study the cohesive devices in these versions. This paper wants to find some similarities and differences in these two versions and may also give some implications for Chinese-English translation.

Index Terms—English versions of *The Analects of Confucius*, cohesive devices, comparative study, corpus.

I. INTRODUCTION

Translation is not simply a process of translating one language into another. The process of translation is actually a process of understanding, analyzing and translating the original text. In this complex activity, translators must face and deal with not only individual words or sentences, but also those interrelated and interdependent words and sentences according to certain rules. Together, these interrelated contents express specific semantic meanings in a certain linguistic environment. Therefore, in the process of translation, translators need to have a certain sense of the text and use appropriate methods to connect words and sentences in order to achieve the accuracy and smoothness of the translation. Under such circumstances, cohesion theory has become an important theory for translators to understand the semantics of the source text, organize and integrate the translation of content and transform it.

As an important theory of discourse analysis and translation, cohesion theory must understand the meaning of discourse. In linguistics, discourse is a higher level of language unit than sentence, and the characteristics of discourse as a whole are also called textuality. This concept was first put forward by Halliday. Neubert put forward a more complete classification of this theory based on the previous researchers, which includes seven important aspects, namely coherence, cohesion, intentionality, acceptability, informativity, situationality and intertextuality. These seven characteristics are interrelated, enabling the text to fully express its meaning and completing its communicative activities. In these seven characteristics, cohesion and coherence are two different properties, but they are closely related to each other. Cohesion is the general term for the method of connecting words and sentences through grammar or vocabulary. It is a visible linguistic phenomenon on the surface of a text and a tangible network. Cohesion means connect dispersed words and sentences into a unified whole, making the content of the article fluent and concrete. Coherence is the internal logical structure of a text, the phenomenon at the bottom of the text, or the invisible network of the text. The definition of cohesion and coherence is sufficient to show their important role in discourse and text translations.

At the same time, the cohesive devices in English and Chinese have similarities, but also have some differences because of the different ways of thinking, cultural background and ways of expressions. So how does this difference reflect in the translations? Will there be some differences in the use of cohesive devices between English and Chinese translators with different backgrounds of mother tongue in C-E translation? This paper intends to make a comparative study of cohesive devices in the English versions of *The Analects of Confucius* by translators with different language background in order to find out some similarities and differences, and to explore the possible reasons for such or other rhetorical preferences.

II. TEXT INTRODUCTION AND RESEARCH METHODS

A. Text Introduction

This study is mainly two English versions of *The Analects of Confucius*. According to the purposes mentioned before, the author selects two English versions of *The Analects of Confucius* from two translators with different cultural and language backgrounds and then makes a small corpus, in order to make an analysis and evaluation of the texts on this basis.

As one of the ancient Chinese *Four Books*, *The Analects of*

Confucius plays an important role in the history of Chinese literature and the construction of Chinese thoughts and values. It is a book that records Confucius' words and actions in his lifetime, but actually it wasn't written by Confucius himself. It was collected and published by Confucius' disciples and reincarnated disciples.

The Analects of Confucius was written in the early Warring States Period. In the long process of dissemination of this book, many scholars from different times had been constantly polishing and adding something to the original one. Although many people have contributed their wisdom to this book, it can be generally ascertained that the authors of this book belong to the class of Shi in ancient times, which means the scholars at that time, which determines that it is a book of high literary and artistic quality. Zhao Pu, a prime minister of the Northern Song Dynasty, once said that one man could govern the whole world after reading just half of *The Analects of Confucius*. It can be easily found out that the status of this work is especially high and it's also greatly influence the Chinese culture. At the same time, as mentioned above, it has been regulating and restricting the moral behavior of the entire Chinese nation, which has great influence and wide spread. It can be regarded as a model of Chinese moral.

Because of its own importance, thousands of scholars and translators have made a lot of effort in translating and researching on it. One of the translators who could not be neglected to be mentioned is James Legge, an English missionary, whose translation version will be discussed in this paper. His hard-work and influence on the English translation of *The Analects of Confucius* is so great just like a milestone in the history of the field of literature. His translation has also greatly influence the translation practices of later scholars. He has provided a series of reliable examples and even templates for its works of translation. However, one thing has to be admitted is that Legge's original intention of translating this work was to preach. At that time, he regarded the translation of Chinese classic literatures into English as an important part of his missionary work, so his version is full of the Christian color which could hardly be ignored. Ku Hungming thought that Legge's translation was too rigid and pedantic, and did not really correctly shape the image of Confucius and Chinese culture.

As for the other important translator in the history of translating *The Analects of Confucius*, Ku Hungming was the man with the background and education of Chinese culture and language. However, he was proficient in Western science and linguistics. In the late Qing Dynasty, he was always enthusiastic about promoting the traditional Chinese culture to the western countries. In this way, he was not only a scholar, but also a patriotic warrior. The main features of Ku Hungming's translation are different from those of Legge. As mentioned above, Legge was careful to follow the original text and tried his best to use the sentence pattern of the original one. Ku Hungming added and deleted parts of the content in some parts so that readers in western countries could better understand the original meanings. Ku expected to be able to translate not only the meaning but also the style of the original text.

B. Research Methods and Procedures

Based on the corpus, this paper attempts to explore the similarities and differences in the use of cohesive devices between two English translators of *The Analects of Confucius* through quantitative and qualitative research. The whole part of this research will be classified into four steps. The first step is to prepare the corpus. The author of this paper finds two versions online and then converts them into text format files in order to build a small corpus. Then the author uses TagAnt 1.2.0 to annotate the texts to make the further research more accuracy. The third step is to use AntConc 3.5.8 to retrieve and count the numbers of each cohesive device in the corpus. The final step is to correct the mistakes in the research results and discuss the reasons of their different translation and whether the different cultural backgrounds affect the translators' use of cohesive devices in the translated text and the influence of the source text on the use of cohesive devices in the translated text.

III. A Contrastive Analysis of Cohesive Devices in English Versions of *The Analects of Confucius*

A. Cohesive Devices

Halliday and Hasan [1] first give a clear classification of cohesion theory in their publication, which is thought as the basis for further study. In this book, they divide cohesion into two categories that are grammatical cohesion and lexical cohesion. Hoey [2] believes that cohesion is the way to connect sentences of the whole discourse through some words or grammatical ways and what Halliday and Hasan mention in their book are mainly lexical cohesion. The visible forms of grammatical cohesion and lexical cohesion are called cohesive devices. The former one, grammatical cohesive device, can be divided in four detailed aspects: reference, substitution, ellipsis, conjunction. While Halliday [1] separates lexical cohesion devices into two specific parts that are reiteration and collocation. And every detailed aspect above can still be divided into smaller parts in the application of analyzing the target text. This paper does this research under the guidance of cohesion theory and devices put forward by Halliday and Hasan. It's not only because their accomplishments in this field but also due to their limpid and explicit classification. In accordance with their classification, in grammatical cohesive device, reference can be classified into personal reference, demonstrative reference and comparative reference. Reference which is what people say or write that mentions other people or things, refers to the mutual expression of two subjects. Substitution is the way to replace one linguistic item, which is also the relationship between grammatical and lexical level. The aim to use substitution in the discourse is to make its expression brief but exact. It contains nominal substitution, verbal substitution and clausal substitution. Ellipsis is also a kind of substitution. It illustrates the trade-off between compactness and clarity. It is the way to omit the repeated component in the sentences, which is also one of the best ways to avoid repetition. It can divide into three parts: nominal ellipsis, verbal ellipsis and clausal ellipsis. The final category of grammatical is conjunction. It is not only an anaphoric, but also expresses some meanings that contain the expression of

other components in the following sentences. In order to facilitate discourse analysis, Halliday and Hasan categorize conjunction, which includes additive conjunction, adversative conjunction, causal conjunction and temporal conjunction. Another main category of cohesive devices is lexical cohesion. Lexical cohesion refers to use works to achieve the aim to make discourse coherence. It can further divide into synonymy and collocation. Halliday [3] makes a explanation of lexical cohesion and further separates it into reiteration and collocation.

Here, the author is not going into make further research about lexical cohesion because of the limited space. This paper will put more attention on grammatical cohesion to do this research. According to the statistics, Legge's version has 27511 tokens in total, while Ku's version has 32718 tokens.

B. Analysis and Discussion of Reference

Reference as a grammatical cohesive device plays a really important role in the cohesion theory. There exits three types of reference, those are personal reference, demonstrative reference and comparative reference. Personal reference is generally achieved its goal through using personal pronoun, possessive determinative, and possessive pronoun. Among them, personal reference is the most widely used and has the most practical and theoretical value in translation studies. As for personal pronouns, it can refer to three kinds: first person, second person and third person.

TABLE I: FREQUENCY AND PERCENTAGE OF PERSONAL REFERENCE

	Legge's translation		Ku's translation	
Personal reference	frequency	percentage	frequency	percentage
First person reference(I, me, my ,mine, we, our, ours)	555	2.02%	598	1.82%
Second person reference(you, your, yours)	218	0.79%	422	1.29%
Third person reference(she, her, hers, he, his, him, it, its, they, them, their, theirs)	1760	6.40%	1739	5.32%

From the Table I, it's easy to find out that both two translators would like to use the third person reference in their translating practices. The original text doesn't have any specific subjects or just uses people or person as the subject in each conversation. It's more appropriate for translators to use the third person reference for better translating of the original one. However, Ku Hungming used more second person reference than Legge. Just because this work is more about a book to teach people some tips to conduct them in a society, Ku used more second person reference close the distance between authors and readers and better conveyed the original meaning. Here is an example.

由子曰："信近於义，言可复也。恭近於礼，远耻辱也。因不失其亲，亦可宗也。"

Legge: The philosopher You said, "When agreements are made according to what is right, what is spoken can be made good. When respect is shown according to what is proper, one keeps far from shame and disgrace. When the parties upon whom a man leans are proper persons to be intimate with, he can make them his guides and masters." [4]

Ku: A disciple of Confucius remarked, "If you make promises within the bounds of what is right, you will be able to keep your word. If you confine earnestness within the bounds of judgment and good taste, you will keep out of discomfiture and insult. If you make friends of those with whom you ought to, you will be able to depend upon them." [5]

As we can see in this example the original text and Legge's version don't appear the exact subject or any personal references, while Ku's version has. But the original sentence has contained the personal references and subjects in its meaning. Using the word "you" can make the sentences more logical and connect better. Readers can clear and easily understand what the author means.

The second reference is demonstrative reference. It refers to use the words like this/these, that/those, the and so on. It is the reference meaning of the location in the space and time. So "this", "these", "now" and "here" are words used to show the proximity to the readers, while "that", "those", "there" and "then" are used to imply distance from the readers, "the" is used as a neutral one.

TABLE II: FREQUENCY AND PERCENTAGE OF DEMONSTRATIVE REFERENCE

	Legge's translation		Ku's translation	
Demonstrative reference	frequency	percentage	frequency	percentage
Distance from the readers(that, those, there, then)	234	0.85%	364	1.11%
Neutral proximity(the)	1763	6.40%	1713	5.24%
Proximity to the readers(this, these, now, here)	154	0.56%	154	0.47%

As we can see in the Table II, in these two versions the translators used more words which show the distance from the speaker. But "the" is the single word that is frequently used. It makes readers to get the meanings of the whole book and have some understandings of that ancient time from a more objective point of view. This book is a record of Confucius and his disciples' words in a conversational way, which includes many principles of living in the world and governing the country. Therefore, both of the translated versions use demonstrative reference without emotional tendencies. The usage of the "the" in the translation versions is more appropriate and specific.

TABLE III: FREQUENCY AND PERCENTAGE OF COMPARATIVE REFERENCE

	Legge's translation		Ku's translation	
Comparative reference	frequency	percentage	frequency	percentage
The adjectives of comparative reference(same, different)	27	0.09%	59	0.18%
The adverbs of comparative reference(better, less, such, so, more)	74	0.27%	144	0.44%
The adverbial phrases of comparative reference(less than, more than)	3	0.01%	4	0.01%

Then it's the comparative reference, which is the reference to use comparative words in order to connect the whole text. For example, the words like same, more and better are

comparative reference. The Table III shows the frequency and percentage of comparative references.

The Analects of Confucius records the words of Confucius and his disciples sentence by sentence, which conveys and educates people's principles of doing things and values of being a nice person. The descriptions are plain and direct, with a minority of comparative words and expressions in the original text. Thus these two translators obeyed the construction of the original one and also used less comparative references. But as we can see in the table above, it can be easily found that Ku Hungming used more comparative references that Legge. And this difference may be due to cultural factors in the two different languages.

As can be seen from the statistics and examples above, three kinds of references are used in both two English translated versions, but Legge and Ku Hungming used more personal pronouns. In addition, because of the differences in hypotaxis and parataxis between English and Chinese, there may not appear the exact references in the Chinese version, which may cause differences in different English versions from different translators. As a Chinese translator, Ku Hungming would much more like to add words or expressions in his translation to make the references much clearer in the original text and make his translation more in line with the English readers' reading habits. If the translation of the reference words is redundant in English translations, in order to make the translation more fluent and clear, some of the cohesive words in the original text will also be omitted.

C. Analyses and Discussion of Substitution and Ellipsis

In general, substitution is the way to replace one linguistic item, which is also the relationship between grammatical and lexical level. The meaning of the substituted words can only be realized in context. It cannot exist independently. Therefore, it is a cohesive device existing at the grammatical and lexical levels. If the substitution words do not appear below, but the meaning is not missing, this is ellipsis. In other words, ellipsis is zero substitution.

TABLE IV: FREQUENCY AND PERCENTAGE OF SUBSTITUTION AND ELLIPSIS

Substitution and ellipsis	Legge's translation		Ku's translation	
	frequency	percentage	frequency	percentage
Nominal substitution and ellipsis(one, ones, the same)	113	0.41%	209	0.64%
Verbal substitution and ellipsis (do, does, did, done, doing)	299	1.09%	286	0.87%
Clausal substitution and ellipsis (so, not)	30	0.11%	29	0.09%

The aim to use substitution in the discourse is to make its expression brief but exact. It contains nominal substitution, verbal substitution and clausal substitution. At the same time, ellipsis is also one of the ways to avoid repetition. It also contains three subcategories just as the substitution: nominal ellipsis, verbal ellipsis and clausal ellipsis. The Table IV is the frequency and percentage of substitution and ellipsis.

In these two translated versions, the translators didn't use

many cohesive devices of substitution and ellipsis. And there is no obvious difference in three parts of substitution and ellipsis, while just the percentage of verbal substitution and ellipsis in Legge's version is a little higher and Ku used more a little more nominal substitution and ellipsis. Here is an example.

子贡问君子。子曰: "先行其言，而后从之."

Legge: Zigong asked what constituted the superior man. The Master said, "He acts before he speaks, and afterwards speaks according to his actions." [4]

Ku: A disciple enquired what constituted a wise and good man. Confucius answered," A wise and good man is one who acts before he speaks, and afterwards speaks according to his actions." [5]

In this example, two translators used different methods to translate the first part of Confucius' words. Legge just used direct expressions and made the appearance of the word "he" twice as a connection. Ku's version has a more complete expression through using the word "one" to refer the man at the beginning of the sentence. Both of these translations are better conveyed the meaning of the Chinese one. Legge's translation is more direct and clear, while Ku's translation is more aesthetic and rhythmic.

In conclusion, according to the research and analysis above, verbal substitution and ellipsis are used most frequent. Among them, verbal substitution/ellipsis and the nominal ones are used more often than clausal substitution/ellipsis. But the frequency of these two cohesive devices is much lower than reference.

D. Analysis and Discussion of Conjunction

Conjunction are the cohesive devices which is different the other grammatical cohesive devices. It is a mean to embody the logical relationships in a discourse by connecting components. So conjunction can also be viewed as the boundary line of grammatical cohesive devices and lexical cohesive device. It helps readers know the semantic connection among sentences and helps them better understand the meaning of conversations. This grammatical cohesive devices can divide into four types，which are additive conjunction, adversative conjunction, clausal conjunction and temporal conjunction.

TABLE V: FREQUENCY AND PERCENTAGE OF CONJUNCTION

Conjunction	Legge's translation		Ku's translation	
	frequency	percentage	frequency	percentage
additive conjunction(and, or, also)	783	2.85%	782	2.39%
adversative conjunction(but, however, though)	165	0.60%	238	0.72%
clausal conjunction (because, so, then)	119	0.43%	149	0.45%
Temporal conjunction(then, next, finally)	53	0.19%	126	0.39%

The additive conjunction uses words like and, or, also and so on. The adversative conjunction uses words like but, however, though and so on. The clausal conjunction uses words like because, so, then and so on. What's more, temporal conjunction uses words like then, next, finally and so on. The following examples are the representations of

using conjunctions.

According to the Table V, it's found that conjunctions are frequently used in both two English versions comparing with the original text. This also confirms the view which some scholars and translators agree, that is, one of the characteristics of translating Chinese classical works into English is that translators should and have to add conjunctions or explanatory words to express the inter-sentence relations if there are no conjunctions between clauses in the original Chinese text. Although there is only a little difference in these two versions, the total percentage of conjunctions in Legge's version is 4.07% which is higher than that in Ku's version of 3.95%. This seems to be related to the translator's different native language and cultural background. As a translator with English taught background, Legge used more explicit cohesive devices like conjunction to show the implicit and logical relationships among sentences in the original text, which obeys the English writing habits and requirements. As a Chinese native translator, Ku Hungming, a well-educated scholar, who was proficient in English, was also influenced by the language and cultural habits of Chinese, which caused the result that the use of conjunction is just a bit of lower. Here is an example.

子曰："弟子，入则孝，出则弟，谨而信，凡爱众，而亲仁。行有馀力，则以学文。"

Legge: The Master said, "A youth, when at home, should be filial, and, abroad, respectful to the elders. He should be earnest and truthful. He should overflow in love to all, and cultivate the friendship of the good. When he has time and opportunity after the performance of these things, he should employ them in polite studies." [4]

Ku: Confucius remarked, "A young man, when at home, should be a good son; when out in the world, a good citizen. He should be circumspect and truthful. He should be in sympathy with all men, but intimate with men of moral character. If he has time and opportunity to spare, after the performance of those duties, he should then employ them in literary pursuits." [5]

In this example, Legge used additive conjunction "and" four times in his translation to reveal the hidden relationship in the original sentence, which fully reflects the linguistic characteristics of English, while Ku Hungming also added additive conjunction "and" and translated adversative conjunction in original sentence "而" into "but", which may not be so accurate as Legge's version. In addition, there are some differences in the numbers of conjunctions they use. Ku's version relies more on the semantic meanings in the sentence to achieve invisible cohesion and retains the short and fluent style of reading the original sentences.

IV. CONCLUSION

Based on the statistical results and analyses of the data above, it can draw the following conclusions. On the one hand, in both two English versions, the translators used many cohesive devices to make their versions much clearer, plainer and easier to understand. Both of them used reference more frequently. But there exit more personal reference in Legge's

translation version and more comparative reference in Ku Hungming's translation version. And there is little difference in substitution and ellipsis. According to conjunction, Legge just used a bit more conjunction than Ku. On the other hand, due to the different language and cultural backgrounds and different strategies of translation adopted by the two translators, there are also some differences in the use of cohesive devices between the two versions.

This paper finds some implications for the further study of discourse analysis and translation practice. First it may help us get a much deeper understanding of cohesive devices of two languages and know the differences of translation practices of translators with opposite cultural and language background. Second it may help translators and researchers think further about the practice of Chinese-English translation. At the same time, the translation of cohesive devices is also related to the quality of the translation version. Accurate translation of cohesive devices in the original text can fully express the meaning and purpose of the original author, which can also bring good reading experience to the target readers and make the target readers get the same reading experience. Therefore, besides the awareness of cohesive devices in translation, translators should first grasp the text as a whole, understand the intention of the author of the original text, base themselves on the reading habits of the target readers and use appropriate translation strategies.

This paper makes a statistical analysis of cohesive devices in two English versions of *The Analects of Confucius*, based on the corpus. The characteristics of the English versions in this respect are presented more intuitively and clearly, which also provides more data support for the study of translation and makes the research conclusions more reliable. However, there are some limitations in this paper. The size of the corpus is small, which only contains two versions. In addition, there may be some subjectivity in explaining the reasons for the similarities and differences between the two versions. These are all aspects that should be focused on in future research.

REFERENCES

[1] M. A. K. Halliday and R. Hasan, *Cohesion in English*, London: Longman, 1976.
[2] Hoey, *Patterns of Lexis in Text*, Oxford: Oxford University Press, 1991.
[3] M. A. K. Halliday and R. Hasan, *Language, Context and Text*, Victoria: Deakin University Press, 1985/1996.
[4] J. Legge, *The Chinese Classics with a Translation, Critical and Exegetical Notes, Prolegomena, and Copious Indexes*, vol. I, Hong Kong: Hong Kong University Press, 1960.
[5] H. M. Ku, *The Discourses and Sayings of Confucius*, Kunming: Yunnan People's Publishing House, 2011.

Yuqiu Hou was born in Baotou, China, September, 1994. She received her bachelor degree of arts in English, Inner Mongolia University, Hohhot, China, 2017.

She is pursing the master degree of arts in Northwestern Polytechnic University in Xi'an, China, in British and American Literature.

Yu Sun was born in Wuhan, China, June, 1971. She was a visiting scholar, University of Maryland, America in 2014. She received her master of arts in English language and literature, Northwestern Polytechnical University, Xi'an, China, 2000.

She is the director in the Arts Education Centre of Northwestern Polytechnical University and also the graduate advisor on literature and translation.

Prof. Sun is also a member of Translators Association of China and the reviewer of China Scholarship Council.