Lecture Notes on Language and Literature (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/langl.2025.080108
ISSN 2523-5869 Vol. 8 Num. 1

A Comparative Study of Translation Quality between Machine Translation Text and Human Translation from the Perspective of Discourse Cohesion Theory—Taking Fond Memories of Peiping as an Example

Zhang Jieru

Guizhou University of Finance and Economics, Guiyang, China
1141827341@qq.com

Keywords: Machine translation, Discourse coherence, Post-editing

Abstract: As the political, economic, and cultural intertwining between countries reaches an unprecedented level of closeness, cooperation in various other fields is also deepening, which implies a significant demand for translation. The rapid development of artificial intelligence has also driven the prosperity of machine translation. Machine translation, with its speed and low cost, has quickly gained popularity in the translation market, becoming an important aid in translation tasks. Machine translation can process large volumes of text in a very short time, saving much time and human resources for translation tasks, but it is not 100% accurate, and the translations produced by machines often require subsequent human proofreading. This article aims to compare machine translation and human translation from the perspective of discourse coherence, using Mr. Lao She's work *Fond Memories of Peiping* as a case study to highlight the differences in discourse coherence between the two translations, identify the issues in machine translation, and improve the quality of machine translation in order to more efficiently promote the development of translation.

1. Introduction

With the acceleration of globalization and the rapid development of information technology, the demand for cross-language communication has become increasingly urgent. In recent years, machine translation technology has made groundbreaking progress, especially with the emergence of neural machine translation technology, which has significantly improved the quality of machine translation and has had a profound impact on the production models of the translation industry [1].

Machine translation is a language generation supported by databases and advanced algorithms. The quality of translation often relates to the quality and applicability of the linguistic collocation corpus resources. "Efficient and larger-scale corpus resources can greatly enhance the efficiency of natural language processing"[2]. In recent years, with the rapid development of artificial intelligence technology, the field of machine translation has achieved significant advancements, especially in neural machine translation (NMT). This technology achieves translation by calculating the neural network relationship between the source and target sentences[3], resulting in more accurate and

fluent translation outputs. However, this model typically works at the sentence level during the translation process, without considering the context of the sentences. Therefore, how to further improve the accuracy, naturalness, and efficiency of machine translation remains a current research focus and challenge.

Against this background, this study selects the prose *Fond Memories of Peiping* from the perspective of discourse cohesion to compare machine translation and human translation, in order to identify the differences between the two types of translation and improve the performance of machine translation.

2. The Discourse Cohesion Theory

The theory of discourse cohesion was proposed by British linguist M.A.K. Halliday in 1962. Later, he and his wife Ruqaiya Hasan published a book titled "Cohesion in English" in 1976, in which they divided cohesion into two main categories: lexical cohesion and grammatical cohesion[4].

Lexical cohesion refers to the relationship between internal vocabulary within a discourse, which can be divided into two main categories: repetition and collocation. Repetition includes recurrence, synonymy, antonymy, hyponymy, and generalization. Collocation creates cohesion by using combinations of words that frequently occur together.

Grammatical cohesion is the external connection between sentences, and common means of grammatical cohesion mainly include reference, substitution, ellipsis, and conjunction[5]. Researchers have conducted a large number of empirical studies based on this theory, further improving and developing the theory of cohesion. For instance, some researchers have explored the relationship between cohesion and coherence, while others have explained this relationship from the perspective of theoretical intersections, such as cognitive psychology, diagrammatic theory, relevance theory, and semantics.

In summary, Halliday's cohesion theory provides an important theoretical framework and tools for discourse analysis, helping us to better understand the coherence and integrity of discourse.

The application of cohesion theory in translation mainly reflects the maintenance and reproduction of semantic coherence in the original text. Translation is a communication between different languages and cultures, which requires the translator to fully understand the semantics and context of the original text and accurately convey this information in the translation. In this process, the use of cohesive devices plays a crucial role.

3. Research Methods and Design

3.1 Corpus Source

This article compares the differences between human translation and machine translation from the perspective of discourse cohesion, thus requiring two versions: machine translation and human translation. The translated text selected is from Mr. Lao She’s work *Xiang Bei Ping* (*Fond Memories of Peiping*). And essay often refers to as "loosely structured yet cohesive in essence." Therefore, special attention should be paid to discourse cohesion in the translation process to ensure the text is coherent and natural. For machine translation, the author decided to use the Deepl translator, which is usually regarded as having higher translation quality compared to other online translation tools. This is mainly due to its use of cutting-edge neural network technology, enabling it to grasp the grammar and semantics of the source language more deeply, thereby producing a smoother and more accurate target language text. The human translation was chosen from Mr. Zhang Peiji's version, who is renowned for his essay translations, making his version a compelling

and valuable reference for comparison.

3.2 Corpus Analysis

The author first organizes and stores the collected machine translations and human translations, then uses the wordlist function in Wordsmith 9.0 to calculate the word frequency of both human and machine translations; the generated word list allows for precise observation of the differences in word distribution between machine translations and human translations, focusing on the top 20 items for analysis. Next, the author manually retrieves various cohesive devices present in both translations for analysis and statistics, and finally generates a list of cohesive words for further analysis. Due to the large number, difficulty in calculation, and minimal impact on the research results, this paper does not include statistics on collocations.

4. Results and Discussion

4.1 Word Frequency Comparative Analysis

The author will use the wordlist function in Wordsmith 9.0 to count the frequency of words appearing in human and DeepL machine translations, and based on this, a word frequency comparison table will be created, focusing on the top 20 vocabulary (as shown in Table 1). This provides an intuitive understanding of the translation situation of both human and machine translations.

Table 1 Word frequency comparison.(top 20)

Human translation				Machine translation			
Order	Word	Frequency	%	Order	Word	Frequency	%
1	the	54	4.34	1	the	41	3.75
2	I	46	3.69	2	and	37	3.39
3	of	39	3.13	3	I	34	3.11
4	and	31	2.49	4	of	31	2.84
5	to	31	2.49	5	a	27	2.47
6	a	30	2.41	6	to	25	2.29
7	in	25	2.01	7	Beijing	19	1.74
8	Beiping	25	2.01	8	in	19	1.74
9	is	20	1.61	9	my	18	1.65
10	as	18	1.45	10	is	15	1.37
11	with	13	1.04	11	that	15	1.37
12	are	12	0.96	12	with	15	1.37
13	my	12	0.96	13	but	12	1.10
14	city	11	0.88	14	it	12	1.10
15	for	11	0.88	15	not	12	1.10
16	it	11	0.88	16	as	11	1.01
17	have	10	0.80	17	its	11	1.01
18	on	10	0.80	18	from	10	0.92
19	can	9	0.72	19	this	9	0.82
20	from	9	0.72	20	for	8	0.73

It can be seen that in both translations, the demonstrative reference "The" is positioned at the beginning, and its frequency is higher in Zhang Peiji's translation, accounting for 4.34%. Similarly,

the demonstrative references "This" and "That" rank in the top 20 in the machine translation, while they do not appear in the top 20 of Zhang Peiji's translation, indicating that the machine translation performs better in using reference methods. The pronoun "it" appears more frequently in the machine translation than in Zhang Peiji's, while the personal pronoun "I" appears less frequently in the machine translation compared to Zhang Peiji's; the frequency of "My" in the machine translation is slightly higher than in Zhang Peiji's. Additionally, the conjunction "And" and the contrasting conjunction "But" also appear more frequently in the machine translation than in Zhang Peiji's. Finally, the term "Beijing" is used less frequently to represent "Beiping" in the machine translation than in Zhang Peiji's translation, but the latter also flexibly uses the alternative term "city," reflecting the flexibility of human translation. This is just a simple analysis of the frequency of word occurrences; the following will specifically analyze the use of cohesive devices in the two translations.

4.2 Comparison and Analysis Results of the Frequency of Cohesion Devices

The author manually indexed the connecting words in the two translations based on the generated vocabulary list, classified their corresponding cohesive devices, and created a comparison table of cohesive devices (Table 2) for analysis.

Table 2 Comparison of cohesive devices between two translations.

cohesive devices		human translation		machine translation	
		word frequency	%	word frequency	%
Grammatical cohesion	reference	104	8.34	112	10.24
	substitution	3	0.24	0	0
	ellipsis	5	0.4	0	0
	conjunction	72	5.78	67	6.14
	total	184	14.76	179	16.38
Lexical cohesion	repetition	55	4.41	42	3.84
	synonym	45	1.44	21	1.9
	hyponymy	11	0.88	5	0.46
	total	111	6.73	68	6.20

From the data presented in Table 2, machine-translated texts display a higher frequency in the usage of grammatical cohesion, reaching 16.38%. Among these, the most frequently used device is reference. In machine translation, the proportion of word used for reference stands at 10.24%, indicating that machine translation performs well in making the article's meaning clear. In Zhang Peiji's translations, the percentages of substitution and ellipsis amount to merely 0.28%, although they are not numerous, they still exist. Conversely, no such means were observed in the machine-translated text, highlighting the professionalism and flexibility of human translator. Regarding the use of conjunctions, there was a relatively high frequency of occurrence in machine translations, accounting for 6.14%, with "And" occurring 37 times. Although the frequency of conjunctions in Zhang Peige's translations is less, the connection methods employed are more diverse, including:(a) Parallel/Consecutive: “and”, “but”; (b) Cause-Effect: “because”; (c) Consecutive/Contrasting: “however”, “nevertheless”; (d) Conditional: “as” and other conjunctions that contribute to enhancing the logical relationship of the article. This suggests that machine translation falls short compared to human translators in context analysis. It often translates word by word, lacking the flexibility in the use of conjunction.

In terms of lexical cohesion, the human translation is more frequent, accounting for 6.73%. It typically employs synonyms, hypernyms and hyponyms, and ellipsis to avoid repetition in English expressions, whereas machine translation lacks the corresponding cohesive devices when encountering repeated words, resulting in poor coherence due to direct repetition of vocabulary. The frequency of synonyms and hypernyms/hyponyms in machine translation is lower than that in Zhang Peiji's translation, which can be attributed to the fact that machine translation often relies on literal translation to reproduce the original text. Human translators consider various factors such as the context of the discourse, the expected audience, and the style of the text when translating. They employ diverse lexical cohesion techniques to enhance the internal logical coherence of the text while maintaining the characteristics of the original work, ensuring that the translation is both faithful and fluent, elegant, allowing the target readers to more easily understand and appreciate the beauty of the article, thus enhancing the fluidity and reading experience of the translation.

From the above data, it shows that machine translation has three problems in discourse coherence: (1) it can use referencing cohesive devices but cannot recognize and apply substitution or omission; (2) it frequently uses “and” as the main linking device and hardly utilizes other conjunctions; (3) apart from 'repetition', it is difficult to see other lexical cohesive devices.

5. Specific Comparison of Cohesion Device Between Human and Machine Translation

5.1 Grammatical Cohesion

5.1.1 Reference

Case 1

MT: By analogy, the little I know is just "my Peking", and my Peking is probably equal to a cow's hair.

HT: It follows that, in contrast with Peiping in its entirety, what little I know about it is probably a mere drop in the ocean[7].

The translation of this sentence clearly shows that Mr. Zhang's version is far superior to machine translation. Phrases like “It follows that” and “a mere drop in the ocean” are better than the machine-translated version. However, today we are focusing on discourse cohesion, so I will temporarily set aside vocabulary issues. The key point of this sentence is “the little I know is just "my Beiping,” which actually includes a comparison between “what I know about Beiping” and “the whole real Beiping”. Chinese is a meaning-concerned language, so the meaning can be understood without being explicitly stated, but English is a form-concerned language. Therefore, in translation, this implicit comparison should be brought to light, namely, “what I know about Beiping is only 'my Beiping.'” Mr. Zhang's translation uses “in its entirety”; “it” refers back to Beiping in his translation, clearly expressing the relationship and cleverly avoiding the repetition of “My Beiping”.

5.1.2 Substitution

Case2

MT: Paris, as far as I can see, is still too busy. Naturally, there are empty and quiet places, but they are too empty.

HT: As far as I know, Paris is too much of a bustling town. It does have quiet open spaces, but they smack of mere expanses of vacancy[7].

The phrase “Paris is too busy” clearly refers to Paris as a bustling city. Mr. Zhang's translation uses "bustling town" to replace the previously mentioned Paris, while the machine translation only

uses the word "busy" to describe that Paris is busy, resulting in a less clear reference compared to Mr. Zhang's translation, which is somewhat inferior.

5.1.3 Ellipsis

Case 3

MT: Paris has been said to be more evenly arranged than London and Rome, but compared to Peking, it is still a bit of a problem. Peking is natural in its artificiality.

HT: Though Paris has a better layout than London or Rome, it nevertheless cannot compare with Peiping, one always finds the natural in the midst of the artificial[7].

This sentence compares the arrangements of Paris with those of London and Rome, omitting “has” after “than London or Rome” in the translation, making it more concise, which machine translation also achieves.

5.1.4 Conjunction

Case 4

MT: There is much in Paris to wear one out, so that coffee and wine are necessary for stimulation; in Peking a mild tea of spiced tablets is sufficient.

HT: While Parisians have to turn to coffee or wine for the relief of boredom caused by so many wearisome places in their city, the mild beverage of jasmine tea will be more than adequate for dwellers of Peiping[7].

From the perspective of cohesion, Mr. Zhang's translation uses the conjunction “while”, which makes the comparison between Beiping and Paris explicit; although the machine translation version also uses the conjunction “so that”, it is a direct translation of the original word “So”, which fails to reflect the contrast in this sentence.

5.2 Lexical Cohesion

5.2.1 Repetition

Case5

MT: This is not only a disappointment to Beiping, but also to myself, for I got my first knowledge and impressions from Beiping, and it is in my blood, and there are many parts of my character and temperament which were given to me by this ancient city.

HT: That is quite a letdown to both Peiping and myself, for it is to this ancient city that I owe what I have within me, including my early knowledge and impressions as well as much of my character and temperament[7].

The word “Beiping” appears twice in this sentence. English tends to avoid repetition, so synonyms, hypernyms, and other alternatives are often used. In Mr. Zhang's translation, the second “Beiping” is replaced with “ancient city” and followed by an attributive clause for clarification, thus avoiding repetition of “ancient city” later in the text. In contrast, the machine translation uses the same word 'Beiping' both times, which feels rigid.

5.2.2 Collocation

Case6

MT: I can't love Shanghai and Tianjin because I have a Beiping in my heart. But I can't say it!

HT: With Peiping possessing my heart, I can never become attached to either Shanghai or Tianjin. I can’t tell why[7].

This sentence, “I can't love Shanghai and Tianjin”, refers to the inability to love either Shanghai or Tianjin. Mr. Zhang's translation cleverly and clearly expresses this meaning by using the “either...or...”, while the machine translation simply used “and”, which is inferior in comparison.

6. Conclusion

This article starts from the perspective of discourse cohesion and compares the differences between machine translation and expert translation. Through deep analysis and comparison, it is found that there are significant differences in the use and effectiveness of cohesive devices between the two.

Machine translation, due to its rule-based and statistical nature, often struggles to fully capture the semantic and contextual information of the original text when dealing with cohesive devices. There are still certain limitations when handling complex cohesive relations, such as substitution and ellipsis. This results in machine-translated texts often being less coherent and natural than those translated by expert translators.

In contrast, expert translators, with their extensive language knowledge and rich translation experience, are able to more accurately understand the semantics and context of the original text when dealing with cohesion devices, and flexibly apply various cohesion techniques in the translation to maintain the coherence and stylistic characteristics of the original.

However, it is worth noting that machine translation and expert translation are not entirely opposed. In fact, with the continuous development of artificial intelligence technology, the combination of machine translation and expert translation has become possible. Future translation research can explore how to combine the advantages of machine translation and expert translation to produce more accurate, natural, and coherent translations.

This article compares the differences between machine translation and expert translation from the perspective of discourse cohesion. Through analysis and comparison, we find significant differences in the use and effectiveness of cohesive devices between the two. Future translation research can further explore how to combine the strengths of both to improve the quality and efficiency of translations.

References

[1] Wang Junsong, Zhuang Zongqian, Wei Yongpeng. Machine Translation Quality Assessment: Methods, Applications, and Prospects [J]. Foreign Language Research, 2024, 40(03):135-144.

[2] Ding Jihong. Research on the Translation of Immigrant Management Empowered by Artificial Intelligence Technology [J]. Journal of Sichuan University of Arts and Science, 2025, 35(01): 102-108.

[3] Guo Mingyang, Zhang Xiaoling, Tang Huiling, et al. Research on the Application of Artificial Intelligence in Machine Translation [J]. Journal of Henan University of Science and Technology: Natural Science Edition, 2021(3): 97-103.

[4] Li Xiaoping. An Inquiry into University English Writing Teaching from the Perspective of Textual Coherence Theory[J]. Modern English, 2024, (08):50-52.

[5] Chen Mingyao; Qiu Hui. A Study on the Cohesion and Coherence of the English Translation of China's Environmental Protection (White Paper) from the Perspective of Discourse[J]. Shanghai Translation, 2009(01):22-26.

[6] Lao She. Thinking of Beiping [J]. Reading, 2022(27):38-39

[7] Zhang Peiji. Selected Modern Chinese Prose in English Translation (Volume 1) [M]. Shanghai: Shanghai Foreign Language Education Press, 2007: 134-137.