

Research on Post-Editing Strategies for Chinese-English Translation Based on ChatGPT

Du Kaiyi

*School of Foreign Languages, Guizhou University of Finance and Economics, Guiyang, China
2517693792@qq.com*

Keywords: Machine translation, Post-editing, ChatGPT, Translation error, Translation strategy

Abstract: The rapid development of artificial intelligence has contributed to the improved speed and efficiency of machine translation. ChatGPT, an advanced language model developed by Open AI, has found widespread application in the field of translation. However, due to the inherent limitations of machines, machine-generated translations still need to be further modified by translators. Today, human-machine collaboration model enjoys high reputation in the translation field. In this context, this article takes translation instances by ChatGPT as the research object, analyzes them from three dimensions of vocabulary, syntax, and discourse, and proposes appropriate post-editing strategies, aiming to further enhance the quality and efficiency of machine translation.

1. Introduction

With the further development of globalization, translation plays a crucial role in cultural exchange and commercial interactions, serving as a bridge between different regions and countries. Nowadays, the upgrading of artificial intelligence has brought great convenience to people's learning and daily lives. ChatGPT, developed by Open AI, is a language model adept at handling various language tasks and has demonstrated outstanding performance in translation. However, despite the widespread recognition of its translation quality, there are still some errors and limitations. In order to further improve the quality of target text, post-editing has garnered attention as a form of human-machine collaboration. This study aims to identify the errors in ChatGPT's translation through case analysis, propose effective post-editing strategies, and evaluate the effectiveness of these strategies.

2. Overview of ChatGPT and Post-Editing

ChatGPT is a large language model based on Transformer, possessing strong generative capabilities and language understanding skills. However, ChatGPT is not a pure machine translation system, as machine translation refers to the automatic translation of text or speech from one natural language to another using computer systems. It can only generate the specified translation based on the given text when it receives translation instructions from human[1]. Nevertheless, due to its strong language processing capabilities, ChatGPT has been widely applied in translation. However,

the complexity of language means that machine translation cannot fully replace human translation, which has made post-editing a focus of many scholars. Post-editing refers to the process of refining and modifying the original output of machine based on specific goals[2]. Cui Qiliang summarizes the practical guidelines for post-editing as follows: determine the post-editing style based on the quality requirements of the translation; identify errors in the machine translation according to the type of source text; establish the post-editing work environment based on the purpose of post-editing; and reasonably set the workload for post-editing on the basis of quality requirements and personnel skills[3]. Moreover, post-editing places higher demands on translators. They often need to possess a solid language foundation, clear logical thinking, extensive professional knowledge, a keen aesthetic sense, and creativity. During post-editing, translators can further polish the machine-generated translation based on principles such as harmonic phonetics, careful word choice, flexible sentence structures, well-crafted organization, appropriate rhetoric, and rich implications[4]. The emergence and application of ChatGPT have significantly promoted the development of post-editing.

3. Overview of Common Errors in Machine Translation

Owing to the significant differences between Chinese and English, as well as the lack of emotions in machines, translations generated by machines can be directly used for some simple texts. However, when it comes to more complex texts, errors often arise, requiring post-editing by translators. Barreiro A et al. identified two major issues affecting sentence translation by machines: “ambiguity” and “cognitive complexity”[5]. Li Ying et al. pointed out that the accuracy of general text translation between Chinese and English is over 90%, which can meet the needs for general text translation. However, translation involves not only the internal structures of sentences but also common knowledge, social knowledge, historical knowledge, psychological states, emotions, and other humanistic elements[6]. When machine encounters these elements, human post-editing becomes particularly important.

4. Case Analysis and Solutions for Post-Editing of Machine Translation Based on ChatGPT

This section will analyze the errors and limitations in the translations generated by ChatGPT from vocabulary, syntax, and discourse, and propose suitable solutions.

4.1 Vocabulary

The translation of words is the foundation of sentence and discourse translation, and mistranslation of words tend to affect the overall quality of translation. ChatGPT often makes some mistakes when translating vocabulary. Different post-editing strategies should be employed based on the type of translation errors. When ChatGPT selects imprecise vocabulary due to misunderstanding of word meanings, the post-editor should refer to English-English dictionaries to choose more appropriate words. If ChatGPT makes translation errors due to literal interpretations, the translator should first decode the vocabulary in the source text and then select suitable translation strategies. In cases of vocabulary redundancy and errors in translating culturally specific terms, the translator should adopt appropriate methods based on realities.

4.1.1 Word Meaning Deviation

Target Text by ChatGPT: The introduction of the unified purchase and sale policy in 1953 was thus shaped by this political mindset.

ChatGPT chose the phrase “political mindset”, failing to understand the meaning of the word

“mindset” accurately. The English dictionary defines “mindset” as “a habitual or characteristic mental attitude that determines how you will interpret and respond to situations”, indicating that “mindset” emphasizes a mental attitude or a thinking tendency. In contrast, “thinking” is defined as “the process of using your mind to consider something carefully”, which emphasizes careful consideration of a matter. Given the context of the source text, the implementation of the unified purchase and sale policy is a result of careful consideration; therefore, choosing “political thinking” is more suitable. In summary, for words with similar meanings, it is advisable to consult their English definitions during post-editing to select more accurate words.

Post-editing: The introduction of the unified purchase and sale policy in 1953 was thus shaped by this political thinking.

4.1.2 Literal Interpretation

Target Text by ChatGPT: Under this premise, the grain issue was not merely an economic matter but more importantly a political one.

The selection of “premise” by ChatGPT indicated that it merely capture the literal meaning while neglecting the context or actual meaning of the source text. Here, the source text aims to convey the meaning: “in this context” or “against this backdrop”. By comparison, translation like “under this premise” sounds more like Chinglish and such problems are pretty common in students writing. Thus, “context” is a more accurate translation than “premise”. At the same time, since the word “context” is used during post-editing, the corresponding phrase should also be adjusted, which means changing the preposition from “under” to “in”. In short, machine tend to convey the literal meaning instead of the implied meaning of the source text because of the absence of subjectivity. Regarding these issues, during following editing, one should first decode the vocabulary and then choose an appropriate translation strategy.

Post-editing: In this context, the grain issue was not merely an economic matter but more importantly a political one.

4.1.3 Lexical Redundancy

Target Text by ChatGPT: Government departments should adopt more policies that are conducive to stabilizing expectations, stabilizing growth, and stabilizing employment.

In ChatGPT’s version, the word “stabilizing” appeared three times, resulting in a rather redundant sentence. Due to the differences between Chinese and English, it is advisable in English to avoid repetition unless for emphasis or rhetorical effect. In contrast, Chinese often employs repetition for grammatical correctness and structural balance[7]. During post-editing, it is recommended to use a combined translation approach to merge the three instances of “stabilizing”, which can make the translation more concise without affecting its meaning. Additionally, since Chinese is a dynamic language that frequently uses verbs, while English is a static language that often employs prepositions, a part-of-speech conversion can be applied here. The verb “stabilizing” is translated into an adjective, and the word “keep” is added, which aligns better with English usage. To sum up, combination is a better choice when faced with such problems.

Post-editing: Government departments should adopt more policies that are conducive to keeping expectations, economic growth, and employment stable.

4.1.4 Mistranslation of Words with Chinese Characteristics

Target Text by ChatGPT: As a populous country, we must adhere to a broad perspective on agriculture and food, always keeping the rice bowl firmly in our own hands.

The phrase “ keeping the rice bowl firmly in our own hands” is a typical expression with Chinese

characteristics. These expressions are very common in Chinese political texts and also pose great challenges for machine translation. ChatGPT chose a literal translation method, conveying only the surface meaning. Actually, this phrase refers to ensuring the stability of food supply. For Chinese people, it is quite easy to grasp the meaning of the phrase. However, the English translation of political texts is targeted at foreign readers, who have difficulty in understanding such terms due to cultural differences. Consequently, translators during post-editing ought to figure out the true meaning of the source text first and then choose proper translation strategies to help foreigners better understand China. Here, the translation “ensure that China’s food supply remains firmly in our own hands” is more appropriate.

Post-editing: As a populous country, we must adopt an all-encompassing approach to agriculture and food and ensure that China’s food supply remains firmly in our own hands.

4.2 Syntax

Chinese emphasizes parataxis, with a loose structure and flexible usage, while English prioritizes hypotaxis, adhering to a strict grammatical system. This creates great differences in expression between Chinese and English, evident not only in word meanings, collocations, and expression convention, but also in sentence structure and logical relationships[8]. Typically, ChatGPT adopts a literal translation approach, which fails to recognize the internal logic of sentences. This often results in inappropriate subject choices and misplaced key information within the sentence. During post-editing, translators should first clarify the internal logic of sentences, choose appropriate subjects based on the source text, and place key information in suitable places.

4.2.1 Unclear Internal Logic

Target Text by ChatGPT: From a domestic perspective, after experiencing the impact of the COVID-19 pandemic for three years, there are numerous challenges to economic recovery and development.

Through a deep analysis, it is clear that the challenges facing the economy are attributed to the three-year COVID-19 pandemic. However, ChatGPT did not capture this causal relationship in its translation, instead rendering it merely as a temporal adverbial phrase. Chinese culture is a high-context culture, which relies heavily on context. Whereas for many English-spoken countries, they belong to low-context culture, which means information should be conveyed clearly and directly. Thus, when translating, the implicit logic in Chinese should be made explicit in English to facilitate understanding for the target readers. In fact, it is a huge challenge for AI. With the help of translators, the addition of “owing to” not only clarifies the logic of the translation but also conveys the original meaning more accurately. Besides, replacing “from a domestic perspective” with “domestically” appears more concise.

Post-editing: Domestically, owing to the impact of a three-year COVID-19 pandemic, there are numerous challenges to economic recovery and growth.

4.2.2 Improper Subject Selection

Target Text by ChatGPT: The plenary session proposed that a high-standard socialist market economy is an important guarantee for Chinese modernization.

It is well known that only humans can propose certain viewpoints or suggestions as the initiators of actions. In the translation, ChatGPT incorrectly chose “the plenary session” as the subject of the entire sentence. Based on the source text, the viewpoint “a high-standard socialist market economy is an important guarantee for Chinese modernization” was presented at this plenary session, with the plenary session functioning as an adverbial component of the sentence. Furthermore, the source

text does not specify who exactly made the proposal. Considering the common use of the passive voice in English, during post-editing, it would be more appropriate to select “it” as a formal subject and use the passive voice. This is more common in English.

Post-editing: In the plenary session, it was proposed that a high-standard socialist market economy is an important guarantee for Chinese modernization.

4.2.3 Inappropriate Location of Key Information

Target Text by ChatGPT: His character was formed precisely in this environment.

In Chinese, key information is often placed at the end of a sentence, typically following an analysis before drawing a conclusion. In contrast, English tends to position important information at the beginning of a sentence, usually presenting ideas from general to specific. The Analysis of the text reveals that the author wants to emphasize the importance of environment, but ChatGPT translated it strictly according to the original word order, which fails to conform to common English usage. During post-editing, the translator took the differences between Chinese and English into account and adjusted the sentence structure by using an emphatic construction to place the key information at the beginning, making it clear to the reader. When faced with such issues, translators can resolve them in the post-editing process by altering the sentence structure.

Post-editing: It was this environment that contributed to his character.

4.3 Discourse

Discourse translation poses a significant challenge for machines. Machine translation typically operates at the level of sentences or paragraphs and cannot integrate context like humans do. From a discourse perspective, translations generated by ChatGPT often exhibit issues such as repetitive word choice and lack of coherence between sentences. To address these issues, translators can adopt strategies during post-editing such as selecting synonyms, choosing appropriate vocabulary based on the specific context, or adding conjunctions to enhance coherence.

4.3.1 Incoherence

Target Text by ChatGPT: The government should deepen cooperation between eastern and western regions and targeted assistance. It should increase support for key counties in the national rural revitalization strategy, establish and improve a regular assistance mechanism for rural low-income populations and underdeveloped regions, to make the results of poverty alleviation more stable and sustainable.

By analyzing the translation by ChatGPT, we found that it ignore the connection between the two sentences. ChatGPT dealt with these two sentences separately, resulting in a version that is incoherent structurally. Halliday and Hasan define cohesion as “the occurrence of cohesion when the interpretation of one element in a text depends on the interpretation of another element in the text[9].” According to the original text, “deepen cooperation, increase support, and establish mechanism” are all human-initiated measures, while “make the results of poverty alleviation more stable and sustainable” is the outcome. Therefore, during post-editing, these three measures were combined in the translation. In view of the characteristics of political texts, the subject “we” was added, which is more common in the official translation of political texts. Moreover, the introduction of the phrase “With these efforts” enable the logical structure of the discourse clearer and the cohesion more coherent. In a word, it is necessary to adjust sentence structure when addressing such problems.

Post-editing: We will deepen east-west cooperation and targeted assistance, provide more support for key counties in the national rural revitalization strategy, and establish sound

mechanisms to provide regular assistance to low-income populations and underdeveloped regions. With these efforts, the achievements in poverty alleviation will be more stable and sustainable.

4.3.2 Monotonous Vocabulary

Target Text by ChatGPT: We need to accelerate the development of our national strategic talent pool, improve the mechanism for discovering and nurturing top-notch innovative talents, establish platforms for cultivating fundamental research talents, and increase support for young scientific and technological talents.

In the translation generated by ChatGPT, the word “talents” appeared four times, resulting in somewhat monotonous vocabulary. It is unavoidable for machine translation, because machines can not exert their own creativity like humans do. In English discourse, there is an emphasis on lexical richness and diversity; therefore, synonyms or alternative expressions should be employed whenever possible without altering the original meaning. During post-editing, the translator translated the four instances of “talents” into “personnel with expertise”, “top-notch innovators”, “talent”, and “scientists and engineers” respectively. This adjustment not only avoids repetition but also aligns better with English expression conventions. Such issues are common in machine translation, and translators should pay special attention to them during post-editing.

Post-editing: We will work faster to build a contingent of personnel with expertise of strategic importance, improve mechanisms for identifying and nurturing top-notch innovators, develop platforms for fostering basic research talents, and enhance support for young scientists and engineers.

5. Conclusion

Generally speaking, although ChatGPT is pretty intelligent, it still cannot achieve 100% accuracy in translation. The combination of machine translation and human translation is necessary to improve both the quality and efficiency of translation. Through the above analysis, this paper put forward some feasible post-editing strategies for reference. At the vocabulary level, attention should be paid to semantic deviations to avoid misinterpretation, ensuring that the translation is concise and clear, particularly when dealing with words with Chinese characteristics. At the syntax level, attention should be given to the internal structure of sentences, their inherent logic, and information focus. At the discourse level, the richness of vocabulary and the coherence between sentences should be emphasized.

References

- [1] Hu Minyu, Yin Jian. (2024). *Exploration of Post-Editing Capabilities in Machine Translation Based on ChatGPT*. *English Square*, 20, 43-46.
- [2] Feng Quangong, Cui Qiliang. (2016). *Post-Editing Research: Focus Analysis and Development Trends*. *Shanghai Journal of Translators*, 6, 67-94.
- [3] Cui Qiliang. (2014). *Discussion on Post-Editing in Machine Translation*. *Chinese Translators Journal*, 6, 68-73.
- [4] Fan Zirui, Yang Wendi. (2024). *Principles and Strategies of Post-Editing in Machine Translation in the era of Human-Machine Collaboration*. *Shanghai Journal of Translators*, 4, 29-34.
- [5] Barreiro, A., Scott, B., Kasper, W. and Kiefer, B. (2011). *Openlogos machine translation: philosophy, model, resources and customization*. *Machine Translation*, 25, 107-126.
- [6] Li Ying, Qu Weiwei, Xiu Shuxin, Xiang Rui. (2024). *Research on Post-Editing Strategies in Machine Translation*. *Journal of Hubei Adult Education Institute*, 2, 96-102.
- [7] Lian Shuneng. (2010). *Contrastive Studies of English and Chinese*. Higher Education Press, Beijing.
- [8] Cai Lijian. (2020). *Twenty Lectures on Chinese-English Translation*. The Commercial Press, Beijing.
- [9] Liu Chendan, Zhao Xiufeng. (2011). *What is Discourse Linguistics?* Shanghai Foreign Language Education Press, Shanghai.